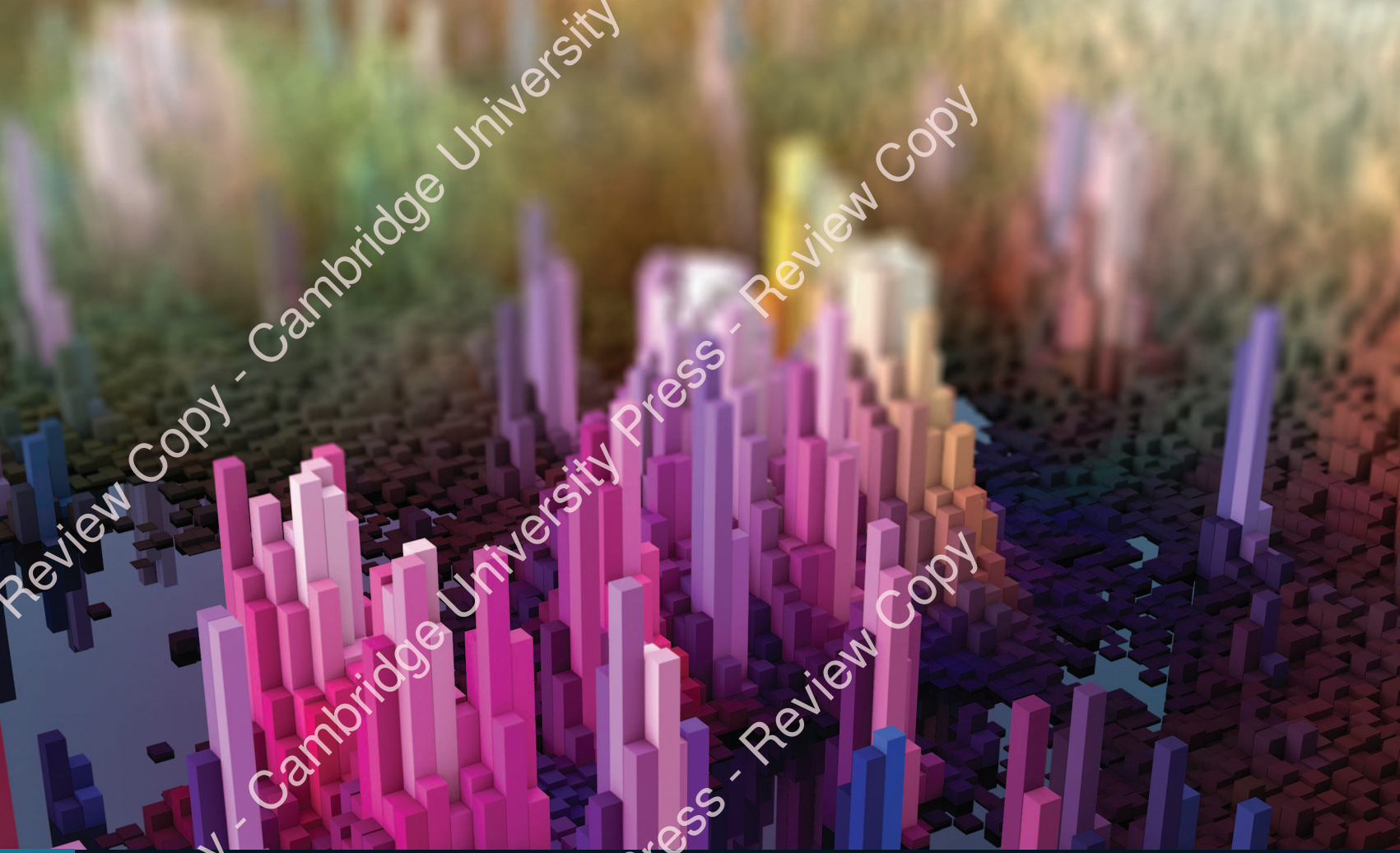
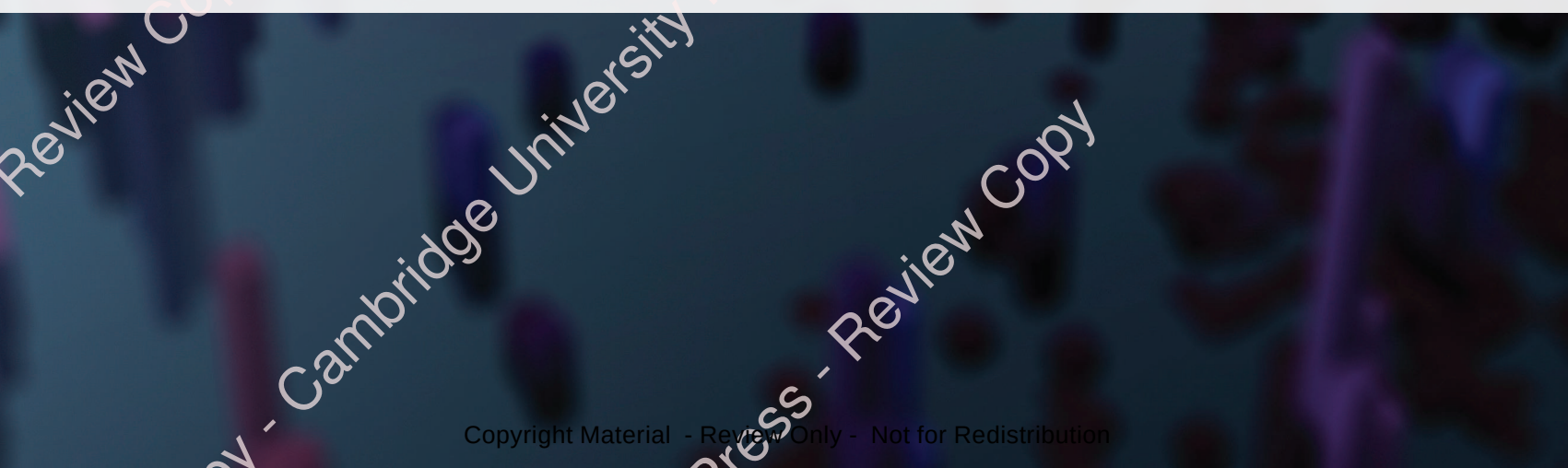


Chapter 6
Estimation

In this chapter you will learn how to:

- calculate unbiased estimates of the population mean and variance from a sample
- formulate hypotheses and carry out a hypothesis test concerning the population mean in cases where the population is normally distributed with known variance or where a large sample is used
- determine and interpret a confidence interval for a population mean in cases where the population is normally distributed with known variance or where a large sample is used
- determine, from a large sample, an approximate confidence interval for a population proportion.



Review Copy - Cambridge University Press - Review Copy
Review Copy - Cambridge University Press - Review Copy
Review Copy - Cambridge University Press - Review Copy
Review Copy - Cambridge University Press - Review Copy
Copyright Material - Review Only - Not for Redistribution

PREREQUISITE KNOWLEDGE

Where it comes from	What you should be able to do	Check your skills
Probability & Statistics 1, Chapters 2 and 3	Calculate the mean and variance from raw and summarised data.	Calculate the mean, variance and standard deviation for the following data sets: **1** $n = 11$ $\Sigma x = 16.5$ $\Sigma x^2 = 25.85$ **2** $n = 8$ $\Sigma x = 434$ $\Sigma x^2 = 26\,630$ **3** 20 24 15 18 16 25 22 **4** 6.5 9.3 13.7 15.1 20.4
Chapter 1	Formulate and carry out hypothesis testing.	State the null and alternative hypotheses and test statistic for the following: **5** $X \sim \mathrm{N}(86, 16)$; sample value 84; two-tailed test at 10% **6** $X \sim \mathrm{N}(54, 3^2)$; sample value 50; one-tailed test at 5% **7** $X \sim \mathrm{N}(18, 3)$; sample value 20; one-tailed test at 1%
Probability & Statistics 1, Chapter 8	Know how to approximate a binomial distribution by a normal distribution.	Express the following as approximate normal distributions: **8** $X \sim \mathrm{B}(42, 0.4)$ **9** $X \sim \mathrm{B}(100, 0.55)$

Why do we study estimation?

Chapter 5 explained that it is not always possible to collect data about every item in a population. There are many practical situations when it is necessary to use a sample to obtain information about a population. For example, an asthma attack may lead to a hospital admission. Sample data allow us to estimate the number of people likely to require a stay in hospital following an asthma attack. Studying the length of the hospital stay will allow us to estimate hospital staffing and other resources. In turn, this allows the hospital to assess its resources and plan for the needs of other patients.

Conservationists study only samples of the population of certain species to make predictions of their numbers. For example, the estimate of the population of mountain gorillas is that there are fewer than 800 left in the world. Snow leopards live in 12 countries in central Asia. Since the start of this century, the estimated number of snow leopards has decreased by about 20%. The actual numbers of snow leopards and mountain gorillas are unknown. These numbers are estimates.

Consider a study that claims two-thirds of adults living in a particular country are overweight. It is unlikely that every adult in that country was weighed; yet the study states they have evidence to justify their claim. That evidence comes from summary statistics from a sample of adults.

The summary statistics calculated from a sample, the sample mean and the sample variance, are used to draw conclusions about the whole population based on the evidence from the sample. These calculated summary statistics, since they only use part of a

Copyright Material - Review Only - Not for Redistribution

population, are estimates. To differentiate between sample statistics and population statistics, the following convention is used:

- Population parameters, such as mean and variance, use Greek letters μ and σ^2, respectively.
- Estimates of population parameters from a sample are written using Roman letters; for example, $\bar{x}$ is the sample mean and s^2 is the sample variance.

Note that the subject you are studying is 'statistics' and confusingly an estimate of a population parameter is called a **statistic**; so estimates of a population's mean and variance are called population statistics.

REWIND

Section 5.2 in the previous chapter explained about the sample mean, $\bar{x}$. This is an estimate for the mean, μ, of a population. The sample mean is an unbiased estimate since the expected value of the sampling distribution of the sample mean is equal to the mean of the population, the parameter it is estimating.

As an example, suppose you wish to find out the average number of fiction books people read each month. You cannot ask the entire population, so instead you ask a sample of the population and work out the average number of fiction books read each month from the sample data. For an unbiased estimate you need to use an unbiased sampling method, such as random sampling, that ensures all members of the population have an equal chance of being selected for the sample; and, of course, you must ask unambiguous questions, making it clear that you are only interested in fiction and not non-fiction books.

6.1 Unbiased estimates of population mean and variance

One objective when taking a sample is to estimate population statistics. A statistic or estimate is a numerical value calculated from a set of data and used in place of an unknown parameter in a population. The bias of an estimate is the difference between the expected value of the estimate and the true value of the parameter. This difference is the **sampling error**. The most efficient estimate is one that is unbiased.

The reliability of an estimate can also depend on the variance of the population. A population with a small variance implies that the data are not widely dispersed and any sample is therefore less likely to be seriously unrepresentative. Conversely, a population with a large variance implies that the data are widely dispersed and so an unrepresentative sample may easily arise.

KEY POINT 6.1

A statistic is an estimate of a given population parameter, calculated from sample data.

A statistic is an unbiased estimate of a given population parameter when the mean of the sampling distribution of that statistic is equal to the parameter being estimated.

If $\hat{U}$ is some statistic derived from a random sample taken from a population, then $\hat{U}$ is an unbiased estimate for U if $\mathrm{E}(\hat{U}) = U$.

The most efficient estimate is one that is unbiased and has the smallest variance.

All the examples presented in Chapter 5 involved a sample from a population with known variance. In practice, if you do not know the population mean, you are unlikely to know the true population variance either.

Copyright Material - Review Only - Not for Redistribution

To explore the sampling distribution of the sample variance, we can return to the example about the spinner numbered 1, 1, 2, 4. In Section 5.2, we found that this distribution has mean 2 and variance 1.5.

To explore the variance as the statistic, for a sample size of 1 we can work out the expectation of the variance $E(V)$.

Sample outcomes	Σx^2	$\bar{x}$	Variance $v = \frac{\Sigma x^2}{1} - \bar{x}^2$	Probability (outcome)
1	1	1	0	$\frac{1}{2}$
2	4	2	0	$\frac{1}{4}$
4	16	4	0	$\frac{1}{4}$

The sample variance, for sample size of 1, $E(V) = 0$.

This is not equal to the variance of the original population, so the sample variance is not an unbiased estimate for the variance.

We do not need to explore the variance for another sample size since a single example that shows the variance is biased is sufficient to prove the point. However, it is worthwhile exploring other sample sizes to see if there is a possible connection between the sample variance and the population variance.

For a sample of size 2, first list all possible sample outcomes, together with the variance and probability of choosing that sample.

Sample outcomes	Σx^2	$\bar{x}$	Variance, $v = \frac{\Sigma x^2}{2} - \bar{x}^2$	Probability (outcome)
1 1	2	1	0	$\frac{4}{16}$
2 2	8	2	0	$\frac{1}{16}$
4 4	32	4	0	$\frac{1}{16}$
1 2	5	$1\frac{1}{2}$	$\frac{1}{4}$	$\frac{4}{16}$
1 4	17	$2\frac{1}{2}$	$2\frac{1}{4}$	$\frac{4}{16}$
2 4	20	3	1	$\frac{2}{16}$

You can check the values for the probabilities of these sample outcomes in Chapter 5, Section 5.2.

We can now draw a probability distribution table for the sample variance, sample size of 2.

v	0	$\frac{1}{4}$	1	$2\frac{1}{4}$
$P(V = v)$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{4}$

Copyright Material - Review Only - Not for Redistribution

Hence, $\mathrm{E}(V) = \left(0 \times \frac{3}{8}\right) + \left(\frac{1}{4} \times \frac{1}{4}\right) + \left(1 \times \frac{1}{8}\right) + \left(2\frac{1}{4} \times \frac{1}{4}\right) = \frac{3}{4}$

Comparing this result with the variance of the original population, we find that $\mathrm{E}(V) = \frac{3}{4}$ and $\sigma^2 = 1\frac{1}{2}$, the variance of the original population.

Notice that $\frac{3}{4} = \frac{1}{2} \times 1\frac{1}{2}$, or $\mathrm{E}(V)\ \frac{n-1}{n} \times \sigma^2$, where n is the sample size.

We need more than just this example to see if this relationship between the original variance and the estimate of variance always holds.

Here are the data for a sample size of 3. You can refer to Chapter 5, Section 5.2 for outcomes and probabilities.

Sample outcome	Σx^2	$\bar{x}$	Variance, $v = \frac{\Sigma x^2}{3} - \bar{x}^2$	Probability (outcome)
1 1 1	3	1	0	$\frac{1}{8}$
1 1 2	6	$\frac{4}{3}$	$\frac{2}{9}$	$\frac{3}{16}$
1 2 2	9	$\frac{5}{3}$	$\frac{2}{9}$	$\frac{3}{32}$
2 2 2	12	2	0	$\frac{1}{64}$
1 1 4	18	2	2	$\frac{3}{16}$
1 2 4	21	$\frac{7}{3}$	$\frac{14}{9}$	$\frac{6}{32}$
2 2 4	24	$\frac{8}{3}$	$\frac{8}{9}$	$\frac{3}{64}$
1 4 4	33	3	2	$\frac{3}{32}$
2 4 4	36	$\frac{10}{3}$	$\frac{8}{9}$	$\frac{3}{64}$
4 4 4	48	4	0	$\frac{1}{64}$

The probability distribution table for the sample variance, sample size of 3, is therefore:

v	0	$\frac{2}{9}$	$\frac{8}{9}$	$\frac{14}{9}$	2
$\mathrm{P}(V = v)$	$\frac{5}{32}$	$\frac{9}{32}$	$\frac{3}{32}$	$\frac{6}{32}$	$\frac{9}{32}$

Hence, $\mathrm{E}(V) = \left(0 \times \frac{5}{32}\right) + \left(\frac{2}{9} \times \frac{9}{32}\right) + \left(\frac{8}{9} \times \frac{3}{32}\right) + \left(\frac{14}{9} \times \frac{6}{32}\right) + \left(2 \times \frac{9}{32}\right) = 1$, and if we use the relationship $\mathrm{E}(V) = \frac{n-1}{n} \times \sigma^2$ for sample size of 3, we find that $\mathrm{E}(V) = \frac{3-1}{3} \times 1\frac{1}{2} = 1$, the same value.

In general, for a population where σ^2 is the population variance, the expectation of sample variance $\mathrm{E}(V)$ is given by:

$$\mathrm{E}(V) = \frac{n-1}{n} \times \sigma^2$$

Copyright Material - Review Only - Not for Redistribution

Using the results we met in Chapter 3, Key point 3.3, this means that $E\left(\frac{nV}{n-1}\right) = \sigma^2$ and $\frac{nV}{n-1} = \frac{n}{n-1}\left(\frac{\Sigma X^2}{n} - \bar{X}^2\right) = \frac{1}{n-1}\left(\Sigma X^2 - n\bar{X}^2\right)$.

KEY POINT 6.2

For sample size n taken from a population, an unbiased estimate of the population mean μ is the sample mean $\bar{x}$.

An unbiased estimate of the population variance σ^2 is:

$$s^2 = \frac{1}{n-1}\left(\Sigma x^2 - n\bar{x}^2\right)$$

TIP

Data may be raw data or summarised data. Use one of the equivalent formulae for variance to suit the information:

$$\frac{1}{n-1}\left(\Sigma x^2 - n\bar{x}^2\right)$$

$$= \frac{1}{n-1}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right)$$

$$= \frac{1}{n-1}\left(\Sigma (x-\bar{x})^2\right)$$

To find an unbiased estimate of variance on your calculator, use σ_{n-1} or s_{n-1}.

WORKED EXAMPLE 6.1

A conservationist wishes to estimate the variance of the numbers of eggs laid by Melodious larks. The following data summarise her results for a sample of 30 Melodious larks' nests (m).

$$\Sigma m^2 = 162,\ \Sigma m = 66$$

Use the data to find an unbiased estimate for the variance of the number of eggs laid by Melodious larks.

Answer

$$\frac{1}{n-1}\left(\Sigma m^2 - \frac{(\Sigma m)^2}{n}\right) = \frac{1}{30-1}\left(162 - \frac{66^2}{30}\right) = 0.579$$

The summarised data suggest using $\frac{1}{n-1}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right)$.

Note that if the question had stated: 'The following data summarise the results of number of eggs laid in 30 nests of Melodious larks. $\Sigma m^2 = 162$, $\Sigma m = 66$. Find the variance.', then the population would be just that group of nests and you would use the variance formula

$$\frac{1}{n}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right) = \frac{1}{30}(162 - 145.2) = 0.56.$$

Note that $\frac{29}{30} \times 0.579\ldots = 0.56$

Copyright Material - Review Only - Not for Redistribution

WORKED EXAMPLE 6.2

A team of conservationists monitoring a tiger population record the number of tiger cubs in a sample of 24 litters. The table shows their findings.

Number of cubs, c	1	2	3	4	> 4
Frequency, f	2	7	12	3	0

Find unbiased estimates for the mean and variance of the number of tiger cubs in the litters.

Answer

$$\bar{c} = \frac{\Sigma fc}{\Sigma f} = \frac{(1\times 2)+(2\times 7)+(3\times 12)+(4\times 3)}{24}$$

$$= \frac{64}{24} = 2\frac{2}{3}$$

$$s^2 = \frac{1}{n-1}\left(\Sigma fc^2 - \frac{(\Sigma fc)^2}{n}\right)$$

$$= \frac{1}{24-1}\left[\left((2\times 1^2)+(7\times 2^2)+(12\times 3^2)+(3\times 4^2)\right) - \frac{64^2}{24}\right]$$

$$= \frac{1}{23}\left[186 - \frac{4096}{24}\right] = \frac{2}{3}$$

Adapt formulae for grouped frequency.

Unbiased estimate for mean is equal to the sample mean.

You can input data into the calculator and use σ_{n-1} or s_{n-1}. Show key values in your working.

EXERCISE 6A

In all of the following questions you are given some data and some descriptive statistics of the data. Your task is to find unbiased estimates of the population mean and variance in each question.

1 Data: the length, x cm, of an electrical component.

$n = 32$, $\Sigma x = 70.4$ and $\Sigma x^2 = 175.56$

2 Data: the time taken, t minutes, in a random sample of dental check-up appointments.

$n = 30$, $\Sigma t = 630$ and $\Sigma t^2 = 13\,770$

3 Data: the yield per plant, in kg rounded to the nearest 100 g, of a random sample of a variety of aubergine plants.

3.5 3.7 4.1 4.4 4.6 4.5 4.5 4.3 4.2

4 Data: the volumes, in ml, for a brand of ice cream in a 750 ml container.

748 751 748 751 745 756 753 760

5 Data: the total mass, x grams, for a random sample of quail eggs.

$n = 16$, $\Sigma x = 128.4$ and $\Sigma x^2 = 1137.6$

6 Data: the total number of faults found in a random sample of 60 silk scarves.

Number of faults per silk scarf	0	1	2	3	4
Number of silk scarves	26	16	7	8	3

Copyright Material - Review Only - Not for Redistribution

7 Data: the time taken, in days, for a random sample of letters posted second class to be delivered.

Number of days for letter to be delivered	1	2	3	4	5
Number of letters	24	32	29	9	6

6.2 Hypothesis testing of the population mean

Sample data are often collected to test a statistical hypothesis about a population. Such a sample, even if it is a random sample, may or may not be representative of the population. The central limit theorem studied in Chapter 5 proves that random sample estimates can be used to make statements about populations without having to assume that the populations have normal distributions. Estimates of the sample mean and sample variance can be calculated from the sample and these estimates can be used to see if they support or reject the null hypothesis. For sample data, $\sqrt{(\text{sample variance})}$; that is, $\frac{\sigma}{\sqrt{n}}$, is referred to as the **standard error**.

We follow the same process as previously used when carrying out a hypothesis test of the population mean. Ideally, we will set up the hypotheses, then collect the sample of data in that order.

REWIND

The steps to carry out a hypothesis test are explained in Chapter 1, but in summary they are:

- Decide whether the situation calls for a one-tailed or two-tailed test.
- State the null and alternative hypotheses.
- Decide on the significance level.
- Calculate the test statistic.
- Compare the calculated probability with the critical value(s).
- Interpret the result in terms of the original claim.

You may also need to consider Type I and Type II errors:

- A Type I error occurs when a true null hypothesis is rejected.
- A Type II error occurs when a false null hypothesis is accepted.

Hypothesis test of a population mean from a normal population with known variance

KEY POINT 6.3

If the population mean is unknown, but the population variance is known, sample data can be used to carry out a hypothesis test that the population mean has a particular value, as follows:

For a sample size n drawn from a normal distribution with known variance, σ^2, and sample mean $\bar{x}$, the test statistic is:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Copyright Material - Review Only - Not for Redistribution

WORKED EXAMPLE 6.3

The masses of cucumbers grown at a smallholding are normally distributed with mean 310 g and standard deviation 22 g. Producers of a new plant food claim that its use increases the masses of cucumbers. To test this claim, some cucumber plants are grown using the new plant food and a random sample of 40 cucumbers from these plants are selected and weighed. The mean mass of these cucumbers is 316 g.

Assuming the standard deviation of the masses of the sample is the same as the standard deviation of the population, test the claim at a 5% level of significance.

Answer 1

Let $\bar{X}$ be the mean mass of cucumbers.

Then $\bar{X} \sim \text{N}\left(310, \frac{22^2}{40}\right)$.

H_0: $\mu = 310$

H_1: $\mu > 310$

One-tailed test at 5% level of significance

$$\text{P}(\bar{X} > 316) \approx \text{P}\left(z > \frac{316 - 310}{\frac{22}{\sqrt{40}}}\right)$$

$$= 1 - \Phi(1.725)$$

$$= 0.0423 \text{ or } 4.23\%.$$

$4.23\% < 5\%$, so the masses are in the critical region.
We reject H_0, because there is some evidence to support the plant food producer's claim.

First set up the test.

Use a one-tailed test, as you are looking for an increase in weight.

Calculate the test statistic using $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ and compare to significance level 5%.

Comment on your result in the context of the question.

Answer 2

Alternative way of writing the solution:

H_0: $\mu = 310$

H_1: $\mu > 310$

One-tailed test at 5% level of significance, critical value z is 1.645.

$z = \frac{316 - 310}{\frac{22}{\sqrt{40}}} = 1.725$ and $1.725 > 1.645$.

So reject H_0 and accept H_1.
There is some evidence to accept the plant food producer's claim.

An alternative approach is to compare the test statistic z with the critical value from tables.

Calculate the test statistic.

Compare it with the critical value, which is $z = \Phi^{-1}(0.95) = 1.645$.

Comment on your result in the context of the question.

WORKED EXAMPLE 6.4

The burn time, in minutes, for a certain brand of candle is modelled by a normal distribution with standard deviation 5.7. The manufacturer claims that the mean is 250 minutes. Lanfen randomly selects ten of these candles and finds that their burn times in minutes are as follows:

245 247 236 255 250 239 241 252 251 243

Stating any assumptions you make, investigate at the 5% level of significance whether the manufacturer's claim is valid.

Copyright Material - Review Only - Not for Redistribution

Answer

Assumptions:

Random sample chosen.

Standard deviation of the sample the same as the population.

Mean $= \frac{1}{10}(245 + 247 + 236 + 255 + 250 + 239 + 241 + 252 + 251 + 243) = \frac{2459}{10} = 245.9$

H_0: $\mu = 250$

H_1: $\mu \neq 250$

Two-tailed test at 5% significance,

critical value z is -1.96.

$\overline{X} \sim N\left(250, \frac{5.7^2}{10}\right)$

$z = \frac{245.9 - 250}{\frac{5.7}{\sqrt{10}}} = -2.275$

$\Phi(-2.275) = 1 - 0.9886 = 0.0114 < 2.5\%$ or

$-2.275 < -1.96$

Reject H_0. There is sufficient evidence to doubt the manufacturer's claim.

The assumptions are the conditions that allow you to use a sample to investigate the claim.

First find the mean of the sample.

State the null and alternative hypotheses.

This is a two-tailed test, as Lanfen is not investigating whether the claim is only too high or only too low.

Use tables to find the critical value, which is $z = \Phi^{-1}(-0.975) = -1.96$.

Calculate the test statistic.

Compare the test statistic with the critical value.

Comment on your result in the context of the question.

Hypothesis test of population mean using a large sample

It is possible to carry out a hypothesis test of a population mean when the population variance is unknown. Provided the sample is large, we follow the same process as for a hypothesis test of a population mean from a normal population with known variance. For the variance, we use s^2, an unbiased estimate of the population variance, where

$$s^2 = \frac{1}{n-1}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right).$$

KEY POINT 6.4

If the population mean and population variance are unknown, sample data can be used to conduct a hypothesis test that the population mean has a particular value, as follows:

For a large sample size n drawn with unknown variance and sample mean $\overline{x}$, the test statistic is:

$$z = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}}$$

where $s^2 = \frac{1}{n-1}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right)$.

Copyright Material - Review Only - Not for Redistribution

WORKED EXAMPLE 6.5

A researcher believes that students underestimate how long 1 minute is. To test his belief, 42 students are chosen at random. Each student, in turn, closes their eyes and estimates 1 minute. The results for their times, x seconds, are summarised as follows:

$$\Sigma x = 2471 \text{ and } \Sigma x^2 = 146\,801$$

Investigate at the 10% level of significance if there is any evidence to support the researcher's claim. What advice would you give to the researcher based on your findings?

Answer

H_0: $\mu = 60$

H_1: $\mu < 60$

One-tailed test at 10% level of significance, critical value z is -1.282.

$$\bar{x} = \frac{\Sigma x}{n} = \frac{2471}{42} = 58.83$$

$$s^2 = \frac{1}{42-1}\left(146\,801 - \frac{2471^2}{42}\right) = 34.73$$

$$z = \frac{58.83 - 60}{\sqrt{\frac{34.73}{42}}} = -1.287$$

$-1.287 < -1.282$

Reject H_0 and accept H_1. There is evidence to support the researcher's claim.

The value of the test statistic and the critical value are very close; advise the researcher to do more tests.

First, state the null and alternative hypotheses. Use 1 minute = 60 seconds.

Decide whether a one-tailed or two-tailed test is appropriate and find the critical value using tables.

Find unbiased estimates for the mean and variance.

Use $\frac{1}{n-1}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right)$ to find an unbiased estimate for the variance.

Next, calculate the test statistic.

Compare the test statistic with the critical value and comment in context of the question.

Advise the researcher of your findings, always in context of the original problem.

EXERCISE 6B

PS **1** The manufacturer of a 'fast-acting pain relief tablet' claims that the time taken for its tablet to work follows a normal distribution with mean 18.4 minutes and variance 3.6^2 minutes^2. Tyler claims that the tablets do not work that quickly. To test the claim, a random sample of 40 people record the time taken for the tablet to work. The mean time for this sample is 19.7 minutes.

Assuming the sample and population variances are the same, carry out an appropriate hypothesis test at the 1% level of significance.

PS **2** IQ test scores are normally distributed and are designed to have a mean score of 100. Anna believes the mean is higher than 100. A random sample of 180 people's IQ test scores, x, are summarised as follows.

$$\Sigma x = 18\,432 \text{ and } \Sigma x^2 = 1\,926\,709.4$$

a Carry out an appropriate hypothesis test at the 2% level of significance.

b Anna then discovers that the IQ test is also designed to have a variance of 15^2. A random sample of six people take the test and their IQ test scores are:

103 109 112 96 100 104

Copyright Material - Review Only - Not for Redistribution

To test Anna's belief, carry out an appropriate hypothesis test, using just the random sample of six people, at the 2% level of significance.

c Comment, with reasons, on the reliability of your answers to parts **a** and **b**.

PS **3** The mass of pesto dispensed by a machine to fill a jar is a normally distributed random variable with mean 380 g. The variance of the mass, in grams^2, of the pesto in the jars is 6.4. Each week a check is made to see that the mean mass dispensed by the machine has not significantly reduced. One particular week a sample of ten jars is checked. The mean mass of pesto in these jars is 378.7 g. Carry out an appropriate hypothesis test at the 5% level of significance, stating any assumptions you have made.

PS **4** The average mass of large eggs is 68 g. The variance of the masses, in grams^2, of large eggs is 1.7^2. A farm shop sells large eggs singly. A customer claims that the eggs are underweight. To test the claim, a random sample of large eggs is weighed. Their masses, in grams, are as follows:

68 65 59 72 65 60 71 73

Carry out an appropriate hypothesis test at the 1% level of significance, stating any assumption(s) you have made.

PS **5** A machine dispenses ice cream into a cone. The amount dispensed follows a normal distribution with mean 80 ml and the variance of the amount of ice cream dispensed, in ml^2, is 9. A consumer complains that the amount is too low. To check whether the machine is dispensing the correct amount, a sample of six cones is checked. The volumes in ml are as follows:

82 72 75 80 76 80

Carry out an appropriate hypothesis test at the 5% level of significance, stating any assumption(s) you have made.

PS **6** A shop sells 2 kg bags of potatoes. A quality control inspection checks the masses of 80 randomly chosen bags. Their masses, x, are summarised as follows:

$\sum x = 158.14$ and $\sum x^2 = 314.094$

Assuming the masses of the bags of potatoes are normally distributed, investigate at the 5% level of significance whether there is any evidence that the bags are underweight.

PS **7** A manufacturer claims its light bulbs last for an average of 2000 hours. A random sample of 42 light bulbs is tested. The lengths of time the light bulbs last, t hours, are summarised as follows:

$\sum t = 83\,895$ and $\sum x^2 = 167\,589\,883.6$

Test the manufacturer's claim at the 10% level of significance, stating any assumptions you have made.

6.3 Confidence intervals for population mean

When a hypothesis test reveals statistically significant results, the results are applicable to the sample. Often we use the results as if they apply to the population. However, we cannot be certain that the sample is actually representative of the population.

The hypothesis tests studied so far, in Chapter 1, Chapter 2 and earlier in this chapter, involve a single parameter, the population mean, from a sample of data. To allow for the issue that the sample may or may not be representative of the whole population, sample data can also be used to construct an interval that specifies the limits within which it is likely that the population mean will lie. This interval is a **confidence interval (CI)**.
A confidence interval for a parameter is constructed at a P% level of confidence such that if the same population is sampled many times and each time an interval estimate is found, the true population parameter will occur in P% of those intervals.

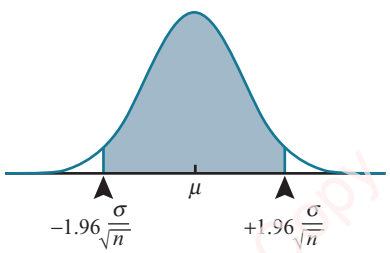
Copyright Material - Review Only - Not for Redistribution

It is possible to construct one-sided or two-sided confidence intervals. However, we will consider only symmetrical two-sided intervals.

A 95% confidence interval is the range of values in which we can be 95% confident that the true mean lies. If that interval is from a to b, then: $P(a < \text{true mean} < b) = 0.95$. The central 95% of the sample distribution is from the 2.5th to the 97.5th percentile.

For a normal distribution $N(\mu, \sigma^2)$, we find from normal tables that the central 95% lies between –1.96 and +1.96 standard deviations either side of the mean.

For a sample distribution, we use $N\left(\mu, \frac{\sigma^2}{n}\right)$.



Although the sample mean, $\bar{x}$, is an unbiased estimate of the true mean, it is only an estimate. It is not necessarily the true mean.

If we work out sample means for a large number of samples, 95% of the time we would expect the sample mean, $\bar{x}$, to lie within the shaded area; that is:

$$\mu - 1.96\frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + 1.96\frac{\sigma}{\sqrt{n}}$$

which rearranges to give:

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

So to find a 95% confidence interval, use the sample values and work out the interval $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$. An alternative way to write the interval is $\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$.

KEY POINT 6.5

A 95% confidence interval means that 95% of possible sample means lie within the interval. It tells us the probability that the true mean lies within the interval is 0.95, and the probability that the true mean does not lie within the interval is $1 - 0.95 = 0.05$.

Confidence intervals for a population mean from a normal population with known variance

Consider Worked example 6.3. The hypothesis test found that there was evidence to accept the producer's claim that its plant food increases the mass of cucumbers.

The sample data are summarised by sample mean $\bar{x} = 316$ and standard error $\frac{22}{\sqrt{40}} = 3.48$.

A 95% confidence interval for these values is $316 \pm 1.96 \times 3.48$; that is, (309, 323).

This means we can be 95% confident that the true mean lies in this range.

Before using the new plant food, the mean mass of cucumbers was 310 g. This mass just lies within the confidence interval (309, 322), at the lower end. We could conclude that it is possible that the plant food does not increase the mass of the cucumbers and the sample is not representative.

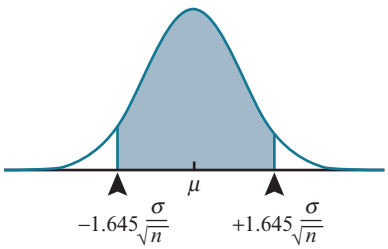
The percentage level chosen for the confidence interval does affect the size of the interval. For example, consider what happens with a 90% confidence interval.

Copyright Material - Review Only - Not for Redistribution

A 90% confidence interval will give a smaller interval.

From normal tables, the central 90% lies within 1.645 standard deviations of the mean.

The 90% confidence interval is $316 \pm 1.645 \times 3.48$; that is, (310, 322).



The lower bound 310 has been rounded from 310.3. Using a 90% confidence interval, the original population mean 310 lies just outside the interval and so you would accept the producer's claim.

What about a 99% confidence interval? Using normal tables, the 99% confidence interval can be calculated as $\mu \pm 2.576 \frac{\sigma}{\sqrt{n}}$, giving $316 \pm 2.576 \times 3.48$; that is, (307, 325).

Compare the confidence intervals, with all values given to the nearest integer: 90% CI = (310, 322), 95% CI = (309, 323) and 99% CI = (307, 325). We can see that the higher the percentage, the more confident we can be that the true mean lies within that interval. However, the higher percentage gives a wider interval, and this means the information we have about the true mean is less precise; that is, there is a greater range of possible values for the true mean.

Sample size also affects the size of a confidence interval.

Consider a population with known standard deviation 15. A random sample $n = 100$ and $\bar{x} = 20$ has standard error $\frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = 1.5$. A 95% confidence interval is then $20 \pm 1.96 \times 1.5$ or (17.1, 22.9).

Let us increase n. If $n = 400$ with $\bar{x} = 20$, then standard error $\frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{400}} = 0.75$ and the 95% confidence interval is narrower, $20 \pm 1.96 \times 0.75$ or (18.5, 21.5).

KEY POINT 6.6

A confidence interval for an unknown population parameter, such as the mean, at a P% confidence level, is an interval constructed so that there is a probability of P% that the interval includes the parameter.

To find the confidence interval for a population mean with known variance σ^2, calculate $\bar{x} \pm k \frac{\sigma}{\sqrt{n}}$, where k is determined by the percentage level of the confidence interval.

% CI	90	95	98	99
k	1.645	1.960	2.326	2.576

The greater the percentage, the more confident we can be that the true parameter lies within the interval.

The greater the percentage, the wider the confidence interval and the less precise we can be about the value of the true parameter.

When choosing the sample size, n, as n increases the standard error $\frac{\sigma}{\sqrt{n}}$ decreases and the resulting confidence interval becomes narrower.

Copyright Material - Review Only - Not for Redistribution

EXPLORE 6.1

As sample size increases, the value of $\frac{\sigma}{\sqrt{n}}$ decreases and so the width of a confidence interval decreases. Why do you think it is not usual practice to use very large samples? Hint: Find the proportional decrease in the width of the confidence interval for different values of n.

Discuss these two questions in relation to the scenarios that follow:

- How large a sample do you actually need?
- What confidence level are you prepared to accept?

Scenario 1: Health officials for a city with population around 40 000 are concerned with the increase in body mass index, BMI, in the population. Would your sample numbers change if the population was, say, 120 000? Explain why or why not.

Scenario 2: A health body wishes to investigate the effectiveness of a new drug treatment. Discuss the possible advantages and disadvantages in combining several different trials of a new drug treatment. (Note that combining the results of many scientific studies is called a meta-analysis.)

WORKED EXAMPLE 6.6

Excessive vegetation in pond water can cause the appearance of unwanted organisms. Over a long period of time it has been found that the number of unwanted organisms in 100 ml of pond water is approximately normally distributed with standard deviation 12. Adam takes six random 100 ml samples of water from his pond. The numbers of unwanted organisms in the samples are 56, 102, 48, 74, 88 and 67.

a Find a 95% confidence interval for the mean number of organisms in 100 ml of the pond water.

b If the mean number of unwanted organisms in 100 ml of pond water is above 80, vegetation should be removed. Use your results to decide whether Adam needs to remove vegetation from his pond. What advice would you give Adam?

Answer

a $\bar{x} = \frac{1}{6}(56 + 102 + 48 + 74 + 88 + 67) = 72.5$

$n = 6, \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{6}} = 4.9$

CI $= 72.5 \pm (1.96 \times 4.9)$ or $(62.9, 82.1)$

First, find the mean of the sample.

Then find the standard error.

Use $\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$ for a 95% confidence interval.

b The upper value of the range of values likely to contain the mean is 82.1. The probability that the mean of 80 organisms lies within this interval is 0.95. Despite the sample mean 72.5 being less than 80, the true mean could be as high as 82.1. Advise Adam to remove some of the vegetation.

Advise Adam to take more samples of pond water.

Use your result to comment in context, to explain and justify the advice.

Copyright Material - Review Only - Not for Redistribution

WORKED EXAMPLE 6.7

The label on a certain packet of sweets states the contents are 100 g. It is known that the standard deviation is 5 g. The mechanism producing these packets of sweets is checked. From a random sample of ten packets, the mean is 103.8 g. Find a 99% confidence interval for the mean contents of the packets of sweets. Use your result to explain whether the mechanism needs adjustment.

Answer

$\bar{x} \pm 2.576 \frac{\sigma}{\sqrt{n}} = 103.8 \pm 2.576 \frac{5}{\sqrt{10}} = 103.8 \pm 4.073$

Use $\bar{x} \pm 2.576 \frac{\sigma}{\sqrt{n}}$.

The CI is $(99.7,\ 107.9)$.

Comment in context. It is important that consumers do not get less than advertised.

The confidence interval tells us that it is possible for the true mean to be below 100 g. The mechanism may need adjustment.

DID YOU KNOW?



Quality control of manufacturing processes is one application of sampling methods. Random samples of the output of a manufacturing process are statistically checked to ensure the product falls within specified limits and consumers of the product get what they pay for. With any product, there can be slight variations in some parameter, such as in the radius of a wheel bolt. Statistical calculations using the distribution of frequent samples, usually chosen automatically, will give information to suggest whether the manufacturing process is working correctly.

Confidence intervals for a population mean using a large sample

To find a confidence interval for a population mean, we rely upon knowing the standard deviation, σ, of the original population. However, since calculating standard deviation involves knowing the mean, it is more likely that the actual value of the standard deviation will be unknown. Instead, we can use the sample data to calculate an unbiased estimate of variance, s^2, and then use s in place of σ to find the confidence interval.

The procedure for finding a confidence interval using an unbiased estimate of standard deviation from a sample gives a reasonably accurate result provided the sample is sufficiently large. How large is sufficiently large? Look back at Explore 6.1. If you compare sample sizes 25 and 400, then since $\sqrt{25} = 5$ and $\sqrt{400} = 20$ you will find that increasing the sample size by 16 times $(16 \times 25 = 400)$ only reduces the margin of error by one-quarter $\left(\frac{1}{\sqrt{25}} \times \frac{1}{4} = \frac{1}{\sqrt{400}}\right)$. 'Large' is not precisely defined; as a general rule, it can be taken to be a sample size of 30 or more.

Copyright Material - Review Only - Not for Redistribution

KEY POINT 6.7

To find the confidence interval for a population mean using a large sample, calculate $\bar{x} \pm k\frac{s}{\sqrt{n}}$, where $s = \sqrt{\frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right)}$ and k is determined by the percentage level of the confidence interval.

WORKED EXAMPLE 6.8

A sample of 60 strawberries is weighed, in grams. The results are summarised as follows:

$$\sum x = 972 \quad \text{and} \quad \sum x^2 = 17\,304.78$$

a Find a 90% confidence interval for the mean mass of the strawberries.

b An α% confidence interval for the population mean, based on this sample, is found to have width of 3.65 grams. Find α.

Answer

a $\bar{x} = \frac{\sum x}{n} = \frac{972}{60} = 16.2$

$s^2 = \frac{1}{60-1}\left(17\,304.78 - \frac{972^2}{60}\right) = 26.413$, so

$s = \sqrt{26.413} = 5.14$

$\bar{x} \pm 1.645\frac{s}{\sqrt{n}} = 16.2 \pm 1.645\frac{5.14}{\sqrt{60}} = 16.2 \pm 1.09$

The CI is $(15.1,\ 17.3)$.

Find unbiased estimates for the mean and variance.

Use $\frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)$ to find an unbiased estimate for the variance.

The sample is sufficiently large to use $\bar{x} \pm 1.645\frac{s}{\sqrt{n}}$.

The confidence interval will be approximate, as the population standard deviation is unknown.

b $2k\frac{5.14}{\sqrt{60}} = 3.65$

$k = \frac{3.65 \times \sqrt{60}}{2 \times 5.14} = 2.75$

For $z = 2.75$, from tables $p = 0.997$.

so $\alpha = 99.4\%$.

The width of the CI is $\pm k\frac{s}{\sqrt{n}}$ or $2k\frac{s}{\sqrt{n}}$. Here you can use the value of s from part **a**.

$1 - p$ gives the percentage in one tail.

$2(1 - p)$ is the percentage in both tails.

$\alpha = 1 - 2(1 - p)$

EXERCISE 6C

For questions 1 and 2 you may refer to the answers to Exercise 6A for unbiased estimates of population mean and variance. Give all confidence limits correct to 3 significant figures.

1 The following data summarise the length, x cm, of an electrical component:

$n = 32$, $\sum x = 70.4$ and $\sum x^2 = 175.56$

Calculate:

a a 98% confidence interval for the population mean

b a 90% confidence interval for the population mean.

Copyright Material - Review Only - Not for Redistribution

2 The following data summarise the time taken, t minutes, in a random sample of dental check-up appointments:

$n = 30$, $\sum t = 630$ and $\sum t^2 = 13\,770$

Calculate:

a a 95% confidence interval for the population mean

b a 98% confidence interval for the population mean.

3 The following data summarise the total mass, x grams of the yield for a random sample of 44 chilli plants.

$\sum x = 842$ and $\sum x^2 = 16\,364$

Calculate:

a a 99% confidence interval for the population mean

b a 95% confidence interval for the population mean.

4 The following data summarise the volume, x litres, for a random sample of bottles of juice.

$n = 68$, $\sum x = 134.14$ and $\sum x^2 = 266.094$

Calculate:

a a 90% confidence interval for the population mean

b a 99% confidence interval for the population mean.

5 The following data summarise the total mass, x grams, for a random sample of quail eggs:

$n = 30$, $\sum x = 254.4$ and $\sum x^2 = 2271.6$

a Calculate a 99% confidence interval for the population mean.

b An α% confidence interval for the population mean, based on this sample, is found to have width of 1.3 grams. Find α.

6 The following data summarise the masses, x kg, of 60 bags of dry pet food.

$\sum x = 117$ and $\sum x^2 = 232.72$

a Calculate unbiased estimates for the population mean and variance.

b Calculate a 98% confidence interval for the population mean.

c An α% confidence interval for the population mean, based on this sample, is found to have width of 0.118 kg. Find α.

7 **a** Explain why the width of a 98% confidence interval for the mean of a standard normal distribution is 4.652.

b The result, X, of testing the breaking strain of a brand of fishing line is a normally distributed random variable with mean μ and variance 2.25. The testers wish to have a 98% confidence interval for μ with a total width less than 1. Find the least number of tests needed.

Copyright Material - Review Only - Not for Redistribution

6.4 Confidence intervals for population proportion

Not every statistical investigation concerns means of samples. Consider, for example, opinion polls. Many organisations carry out opinion polls to gauge voter intentions. Different polls for the same election do not always agree, even when the people chosen are representative samples of the population. When the data have been collected and presented, it is possible to calculate probabilities. We need to question, and statistically calculate, the reliability of their results.

There are a number of situations in which people are required to choose between two options, such as the UK Brexit vote where the options were to either remain or leave the European Union. The following example models an opinion poll for such a situation.

Sofia and Diego are the only two candidates in an election; there is no third option and *everyone* has to vote. Let a vote for Sofia be called a success. In an opinion poll of n people, where r people say they will vote for Sofia, the proportion of successes $\hat{p} = \frac{r}{n}$.

The binomial distribution is a suitable model for this situation since there are only two outcomes, there is a fixed number of people in the poll and each person independently chooses who to vote for.

Let the random variable, X, be the number of people who vote for Sofia. Then $X \sim \mathrm{B}(n, p)$, $\mathrm{E}(X) = np$ and $\mathrm{Var}(X) = np(1-p)$.

Let $\hat{P}$ be the random variable 'the proportion of the sample voting for Sofia'. Then $\hat{P} = \frac{X}{n}$.

Expected value, $\mathrm{E}\left(\hat{P}\right) = \mathrm{E}\left(\frac{X}{n}\right) = \frac{1}{n}\mathrm{E}(X) = \frac{1}{n} \times np = p$, and so $\hat{p}$ is an unbiased estimate for p.

Variance, $\mathrm{Var}\left(\hat{P}\right) = \mathrm{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2}\mathrm{Var}(X) = \frac{1}{n^2} \times np(1-p) = \frac{p(1-p)}{n}$

Before the election, an opinion poll of a random sample of 200 people is conducted. In this opinion poll 108 people say they will vote for Sofia and 92 say they will vote for Diego. With more than half of the people in the sample voting for Sofia, you may conclude that Sofia will win the election. To investigate how reliable this conclusion is we would have to find a confidence interval for the population proportion.

For sufficiently large values of n, such that $np > 5$ and $n(1-p) > 5$, a binomial distribution can be approximated by a normal distribution. So an approximate distribution of the sample proportion is $\mathrm{N}\left(p, \frac{p(1-p)}{n}\right)$.

Confidence intervals for a population proportion are worked out in a similar way to those for the sample mean. We calculate $\hat{p} \pm k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, where k is determined by the percentage level of the confidence interval.

Note that this is an approximate confidence interval since a population proportion has a binomial distribution, which is discrete, whereas the normal distribution is continuous. However, it is not necessary to apply continuity corrections when finding these confidence intervals.

Returning to the opinion poll for Sofia and Diego, for the random sample of 200 people:

Sample proportion, $\hat{p} = \frac{108}{200} = 0.54$

Sample variance $= \frac{\hat{p}(1-\hat{p})}{n} = \frac{0.54 \times (1-0.54)}{200} = 0.001242$

Copyright Material - Review Only - Not for Redistribution

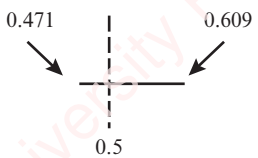
For a 95% confidence interval:

$$\hat{p} \pm k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.54 \pm 1.96 \times \sqrt{0.001242} = 0.54 \pm 0.07$$

So the confidence interval is $(0.471, 0.609)$.

With only two candidates, the winner needs more than 50% of the votes. The question to be resolved is where the confidence interval lies with respect to the 50% value.

The following diagram shows the range of the confidence interval crossing the 50%, or 0.5, mark.



So for this sample, even though more than half said they would vote for Sofia, the confidence interval suggests that the proportion of votes for Sofia could be less than half.

Suppose instead you want to know how many people to poll (i.e. to select for the sample) to find a confidence interval of a given width. You could ask, 'What sample size is needed for an approximate 95% confidence interval for this proportion to have a width of 0.03?'.

To find a confidence interval, we calculate $\pm k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, so the width of the confidence interval is given by $2k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. The question requires the same proportion as the sample, so $\hat{p} = 0.54$.

For a 95% confidence interval we use $k = 1.96$.

$$2 \times 1.96\sqrt{\frac{0.54 \times (1-0.54)}{n}} = 0.03$$

$n = 4240$, to 3 significant figures.

DID YOU KNOW?



A claim by general election opinion polls is that they have a 3% margin of error. In practice, many such polls will have a sample size of 1000.

George Gallup showed the importance of opinion polls when he successfully predicted that Franklin Roosevelt would win the 1936 US presidential election. George Gallup continued to work in the field of public opinion. The work he began studying social, moral and religious opinions continues to this day in over 160 countries.

Today, opinion polls provide research and advice to many large organisations, such as manufacturers of new products, and political parties.

WEB LINK

You can find out more about the guidance of a large UK corporation on conducting and reporting of opinion polls on the BBC website.

EXPLORE 6.2

Find media reports on the results of an opinion poll. Does the report comment on how many people or voters were included in the poll? Does the report comment on the sampling method employed? Use the information given in the report to discuss the reliability of the results in the opinion poll.

Copyright Material - Review Only - Not for Redistribution

KEY POINT 6.8

For a large random sample, size n, an approximate confidence interval for the population proportion, $\hat{p}$, is:

$$\left(\hat{p} - k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

where k is determined by the percentage level of the confidence interval.

WORKED EXAMPLE 6.9

A Sudoku puzzle is classified as 'easy' if more than 70% of the people attempting to solve it do so within 10 minutes, and 'hard' if less than 20% of people take less than 10 minutes to complete it. Otherwise it is classified as 'average'. Of 120 people given a Sudoku puzzle, 87 completed it within 10 minutes.

a Find an approximate 99% confidence interval for the proportion of people completing the puzzle within 10 minutes. Comment on how the Sudoku puzzle should be classified.

b 200 random samples of 120 people were taken and a 99% confidence interval for the proportion was found from each sample. How many of these 200 confidence intervals would be expected to include the true proportion?

Answer

a $\hat{p} = \frac{87}{120} = 0.725$

First, find the sample proportion.

Use 2.58 for a 99% confidence interval.

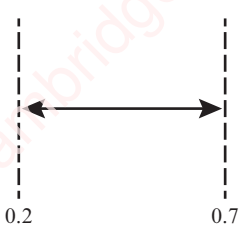
Calculate $\hat{p} \pm k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

Look where the confidence interval lies with respect to the boundaries of interest.

$0.725 \pm 2.58\sqrt{\frac{0.725 \times (1 - 0.725)}{120}}$

$= 0.725 \pm 0.105$

CI = (0.62, 0.83)



The CI crosses the 70% boundary, so although in the sample more than 70% of people completed it in less than 10 minutes, the interval suggests there will be some samples where less than 70% complete it in this time. It could be classified as easy or average.

b 198

200×0.99

TIP

Samples have a margin of error. When you find a confidence interval, you need to consider where it lies in relation to the boundary, or boundaries, used as guides for action.

Copyright Material - Review Only - Not for Redistribution

WORKED EXAMPLE 6.10

Apprentices work four days a week and spend one day a week at college. It is proposed that the college day is changed from a Monday to a Friday. The college will consider changing the day if 80% of apprentices are in favour of the change. In a sample of apprentices, how many should be asked to be 90% certain of gaining 80% support that is not more than 5% wrong.

Answer

$\hat{p} = 0.8$

You want the proportion in favour to be 80% or 0.8.

$5\% = 0.05$

Calculate $k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.05$.

Use 1.645 for 90% confidence interval.

$1.645\sqrt{\frac{0.8 \times (1-0.8)}{n}} = 0.05$

$n = \left(\frac{1.645}{0.05}\right)^2 \times 0.8 \times (1-0.8) = 173$

EXERCISE 6D

1 A quality control check of a random sample of 120 pairs of jeans produced at a factory finds that 24 pairs are sub-standard. Calculate the following confidence intervals for the proportion of jeans produced that are sub-standard:

a a 90% confidence interval

b a 98% confidence interval.

2 At a university, a random sample of 250 students is asked if they use a certain social media app. Of the students in the sample, 92 use this social media app. Calculate a 95% confidence interval for the proportion of students at the university who use this social media app.

3 A four-sided spinner has sides coloured red, yellow, green and blue. The probability that the spinner lands on yellow is p. In an experiment, the spinner lands on yellow 18 times out of 80 spins. Find an approximate 99% confidence interval for the value of p.

PS 4 A biased coin flipped 500 times results in tails 272 times.

a Find a 90% confidence interval for the probability of obtaining a tail.

b This experiment is carried out ten times. How many of the confidence intervals would be expected to contain the population proportion of obtaining a tail?

M 5 The proportion of European men who are red-green colour-blind is 8%. How large a sample would need to be selected to be 95% certain that it contains at least this proportion of red-green colour-blind men?

M PS 6 A random sample of 200 bees from a colony is tested to find out how many are infected with *Varroa* mites. Forty bees are found to be infected.

a Calculate a 99% confidence interval for the proportion of the colony infected with *Varroa* mites.

b The colony of bees will collapse and will not survive if 35% or more are infected with *Varroa* mites. Show why it is possible, at the 99% confidence level, that the colony of bees might collapse.

Copyright Material - Review Only - Not for Redistribution

Checklist of learning and understanding

- If U is some statistic derived from a random sample taken from a population, then U is an unbiased estimate for Φ if $\mathrm{E}(U) = \Phi$.
- For sample size n taken from a population, an unbiased estimate of:
  - the population mean μ is the sample mean $\bar{x}$
  - the population variance σ^2 is:

$$s^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right) = \frac{1}{n-1}\left(\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right) = \frac{1}{n-1}\left(\sum (x-\bar{x})^2\right)$$

- To test a hypothesis about a sample mean, $\bar{x}$, for a sample size n drawn from a normal distribution with known variance, σ^2, calculate the test statistic

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}.$$

  - The test statistic, z, can be used to test a hypothesis about a population mean drawn from any population.
  - Where the population variance is unknown, use the unbiased estimate of variance s^2.
- A confidence interval for an unknown population parameter, such as the mean, is an interval constructed so that it has a given probability that it includes the parameter.
- Confidence interval for:
  - population mean with known variance, σ, is $\left(\bar{x} - k\frac{\sigma}{\sqrt{n}},\ \bar{x} + k\frac{\sigma}{\sqrt{n}}\right)$
  - population mean using a large sample is $\left(\bar{x} - k\frac{s}{\sqrt{n}},\ \bar{x} + k\frac{s}{\sqrt{n}}\right)$, where $s = \sqrt{\frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right)}$
  - population proportion, $\hat{p}$, is $\left(\hat{p} - k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$, where k is determined by the percentage level of the confidence interval.

Review Copy - Cambridge University Press - Review Copy
Copyright Material - Review Only - Not for Redistribution

END-OF-CHAPTER REVIEW EXERCISE 6

PS **1** The worldwide proportion of left-handed people is 10%.

a Find a 95% confidence interval for the proportion of left-handed people in a random sample of 200 people from town A. **[3]**

b In town B, there is a greater proportion of left-handed people than there is in town A. From a random sample of 100 people in town B, an α% confidence interval for the proportion, p, of left-handed people is calculated to be (0.113, 0.207).

i Show that the proportion of left-handed people in the sample from town B is 16%. **[2]**

ii Calculate the value of α. **[3]**

PS **2** The label on a jar of jam carries the words 'minimum contents 272 g'.

a Explain why, in practice, the average contents need to be greater than 272 g. **[1]**

b The mass of jam dispensed by a machine used to fill the jars is a normally distributed random variable with mean 276 g. The variance of the mass of the jam, in grams^2, dispensed by the machine is 1.8^2. Each week there is a check to see if the mean mass dispensed by the machine is 276 g. One particular week a sample of eight jars is checked. The mean mass of jam in these jars is 277.7 g. Carry out an appropriate hypothesis test at the 5% level of significance, stating any assumptions you have made. **[5]**

M **3** An employer who is being sued for the wrongful dismissal of an employee is advised that any award paid out will be based on national average earnings for employees of a similar age. A random sample of 120 people is found to have a mean income of \$21 000 with standard error \$710.

a Find a 95% confidence interval for the award. **[3]**

b The employer wants to know the upper limit of the award that is very unlikely to be exceeded. The employer defines 'unlikely' as a probability of 0.001.

i Explain why the required size of the confidence interval is 99.8%. **[1]**

ii Work out the unlikely upper limit of the award, giving your answer to the nearest dollar. **[3]**

M **4** The volume, v ml, of liquid dispensed by a vending machine for a random sample of 60 hot drinks is summarised as follows:

$$\Sigma v = 17\,280 \text{ and } \Sigma v^2 = 5\,015\,000$$

a Find unbiased estimates of the population mean and variance. **[2]**

b Work out a 90% confidence interval for the population mean. **[3]**

M PS **5** The manufacturer of a certain smartphone advertises that the average charging time for the battery is 80 minutes with standard deviation 2.6 minutes. Owners of these smartphones suggest that the time is longer. A random sample of the phones were charged from 0 to 100% and their times, in minutes, are as follows.

88 85 82 77 86 75 80 79

a Investigate at the 5% level of significance whether or not the manufacturer's claim is justified, stating any assumption(s) you have made. **[6]**

b The given length of time a charged smartphone battery will last is normally distributed with mean 24 hours. The variance of the time taken, in minutes, for the smartphone to work is 1. The variance of the length of time time taken, in hours squared, for the battery to last is 1. Sami tests a random sample of five batteries;

Copyright Material - Review Only - Not for Redistribution

the sample mean time is 23.2 hours. Investigate at the 5% level of significance whether the time batteries last is less than the time given, stating any assumption(s) you make. **[5]**

c In a single sample, determine how long the battery could last for, if a Type I error occurs. **[1]**

6 The manufacturer of a tablet computer claims that the mean battery life is 11 hours. A consumer organisation wished to test whether the mean is actually greater than 11 hours. They invited a random sample of members to report the battery life of their tablets. They then calculated the sample mean. Unfortunately a fire destroyed the records of this test except for the following partial document.

Test of the mean batter the tablet	
Sample size, n	
Sample mean (hours)	11.8
Is the result significant at the 5% level?	Yes
Is the result significant at the 2.5% level?	No

Given that the population of battery lives is normally distributed with standard deviation 1.6 hours, find the set of possible values of the sample size, n. **[5]**

Cambri vel aper 73 Q4 November 2016

7 Parcels arriving at a certain office have weights W kg, where the random variable W has mean μ and standard deviation 0.2. The value of μ used to be 2.60, but there is a suspicion that this may no longer be true. In order to test at the 5% significance level whether the value of μ has increased, a random sample of 75 parcels is chosen. You may assume that the standard deviation of W is unchanged.

i The mean weight of the 75 parcels is found to be 2.64 kg. Carry out the test. **[4]**

ii Later another test of the same hypotheses at the 5% significance level, with another random sample of 75 parcels, is carried out. Given that the value of μ is now 2.68, calculate the probability of a Type II error. **[5]**

Cambridge International AS & A Level Mathematics 9709 Paper 73 Q6 November 2015

8 Last year Samir found that the time for his journey to work had mean 45.7 minutes and standard deviation 3.2 minutes. Samir wishes to test whether his average journey time has increased this year. He notes the times, in minutes, for a random sample of 8 journeys this year with the following results.

46.2 41.7 49.2 47.1 47.2 48.4 53.7 45.5

It may be assumed that the population of this year's journey times is normally distributed with standard deviation 3.2 minutes.

i State, with a reason, whether Samir should use a one-tail or a two-tail test. **[2]**

ii Show that there is no evidence at the 5% significance level that Samir's mean journey time has increased. **[5]**

iii State, with a reason, which one of the errors, Type I or Type II, might have been made in carrying out the test in part **ii**. **[2]**

Cambridge International AS & A Level Mathematics 9709 Paper 73 Q6 June 2012

Copyright Material - Review Only - Not for Redistribution

9 The management of a factory thinks that the mean time required to complete a particular task is 22 minutes. The times, in minutes, taken by employees to complete this task have a normal distribution with mean μ and standard deviation 3.5. An employee claims that 22 minutes is not long enough for the task. In order to investigate this claim, the times for a random sample of 12 employees are used to test the null hypothesis $\mu = 22$ against the alternative hypothesis $\mu > 22$ at the 5% significance level.

i Show that the null hypothesis is rejected in favour of the alternative hypothesis if $\bar{x} > 23.7$ (correct to 3 significant figures), where $\bar{x}$ is the sample mean. **[3]**

ii Find the probability of a Type II error given that the actual mean time is 25.8 minutes. **[4]**

Cambridge International AS & A Level Mathematics 9709 Paper 71 Q5 November 2011

10 A doctor wishes to investigate the mean fat content in low-fat burgers. He takes a random sample of 15 burgers and sends them to a laboratory where the mass, in grams, of fat in each burger is determined. The results are as follows.

9 7 8 9 6 11 7 9 8 9 8 10 7 9 9

Assume that the mass, in grams, of fat in low-fat burgers is normally distributed with mean μ and that the population standard deviation is 1.3.

i Calculate a 99% confidence interval for μ. **[4]**

ii Explain whether it was necessary to use the Central Limit Theorem in the calculation in part **i**. **[2]**

iii The manufacturer claims that the mean mass of fat in burgers of this type is 8 g. Use your answer to part **i** to comment on this claim. **[2]**

Cambridge International AS & A Level Mathematics 9709 Paper 72 Q4 June 2011

11 The masses of sweets produced by a machine are normally distributed with mean μ grams and standard deviation 1.0 grams. A random sample of 65 sweets produced by the machine has a mean mass of 29.6 grams.

i Find a 99% confidence interval for μ. **[3]**

The manufacturer claims that the machine produces sweets with a mean mass of 30 grams.

ii Use the confidence interval found in part **i** to draw a conclusion about this claim. **[2]**

iii Another random sample of 65 sweets produced by the machine is taken. This sample gives a 99% confidence interval that leads to a different conclusion from that found in part **ii**. Assuming that the value of μ has not changed, explain how this can be possible. **[1]**

Cambridge International AS & A Level Mathematics 9709 Paper 73 Q3 November 2010

12 A random sample of n people were questioned about their internet use. 87 of them had a high-speed internet connection. A confidence interval for the population proportion having a high-speed internet connection is $0.1129 < p < 0.1771$.

i Write down the mid-point of this confidence interval and hence find the value of n. **[3]**

ii This interval is an α% confidence interval. Find α. **[4]**

Cambridge International AS & A Level Mathematics 9709 Paper 71 Q2 June 2010

Copyright Material - Review Only - Not for Redistribution

13 The masses of packets of cornflakes are normally distributed with standard deviation 11 g. A random sample of 20 packets was weighed and found to have a mean mass of 746 g.

i Test at the 4% significance level whether there is enough evidence to conclude that the population mean mass is less than 750 g. **[4]**

ii Given that the population mean mass actually is 750 g, find the smallest possible sample size, n, for which it is at least 97% certain that the mean mass of the sample exceeds 745 g. **[4]**

Cambridge International AS & A Level Mathematics 9709 Paper 72 Q5 November 2009

Review Copy - Cambridge University Press - Review Copy

Copyright Material - Review Only - Not for Redistribution

CROSS-TOPIC REVIEW EXERCISE 2

M 1 The time to failure, in years, for two types of kettle can be modelled by the continuous random variables X and Y which, respectively, have probability density functions as follows:

$$f(x) = \begin{cases} \frac{x}{8} & 0 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases} \qquad f(y) = \begin{cases} \frac{9}{4y^3} & 1 \leqslant y \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

Show that the probability of failure by time t is the same for both X and Y if t satisfies the equation $t^4 - 18t^2 + 18 = 0$ and verify that this time is just over 1 year. **[7]**

2 A continuous random variable X has probability density function given by:

$$f(x) = \begin{cases} 0.25 & 4 \leqslant x \leqslant 8 \\ 0 & \text{otherwise} \end{cases}$$

a Sketch the graph of $y = f(x)$. **[2]**

b State the mean and use integration to find the variance. **[3]**

c The mean of a random sample of 40 observations of X is denoted by $\bar{X}$. State the approximate distribution of $\bar{X}$, giving its parameters. **[3]**

d Find the value of a, where $P(\bar{X} < a) = 0.9$. **[3]**

M 3 Chakib cycles to college. He models his journey time, T minutes, by the following probability density function:

$$f(t) = \begin{cases} \frac{1}{100}(25 - t) & 10 \leqslant t \leqslant 20 \\ 0 & \text{otherwise} \end{cases}$$

a Work out the mean and variance. **[7]**

Chakib finds that a random sample of 20 of his journey times has mean 12.4 minutes.

b Write down the approximate distribution of the sample mean for a sample of size 20. **[3]**

c Show that Chakib's model is not suitable. **[4]**

PS 4 **i** Give a reason for using a sample rather than the whole population in carrying out a statistical investigation. **[1]**

ii Tennis balls of a certain brand are known to have a mean height of bounce of 64.7 cm, when dropped from a height of 100 cm. A change is made in the manufacturing process and it is required to test whether this change has affected the mean height of bounce. 100 new tennis balls are tested and it is found that their mean height of bounce when dropped from a height of 100 cm is 65.7 cm and the unbiased estimate of the population variance is $15\,\text{cm}^2$.

a Calculate a 95% confidence interval for the population mean. **[3]**

b Use your answer to part **ii a** to explain what conclusion can be drawn about whether the change has affected the mean height of bounce. **[1]**

Cambridge International AS & A Level Mathematics 9709 Paper 72 Q3 June 2016

Copyright Material - Review Only - Not for Redistribution

5 The diameter, in cm, of pistons made in a certain factory is denoted by X, where X is normally distributed with mean μ and variance σ^2. The diameters of a random sample of 100 pistons were measured, with the following results:

$$n = 100 \quad \Sigma x = 208.7 \quad \Sigma x^2 = 435.57$$

i Calculate unbiased estimates of μ and σ^2. **[3]**

The pistons are designed to fit into cylinders. The internal diameter, in cm, of the cylinders is denoted by Y, where Y has an independent normal distribution with mean 2.12 and variance 0.000144. A piston will not fit into a cylinder if $Y - X < 0.01$.

ii Using your answers to part **i**, find the probability that a randomly chosen piston will not fit into a randomly chosen cylinder. **[6]**

Cambridge International AS & A Level Mathematics 9709 Paper 73 Q7 November 2015

6 The marks, x, of a random sample of 50 students in a test were summarised as follows:

$$n = 50 \quad \Sigma x = 1508 \quad \Sigma x^2 = 51\,825$$

i Calculate unbiased estimates of the population mean and variance. **[3]**

ii Each student's mark is scaled using the formula $y = 1.5x + 10$. Find estimates of the population mean and variance of the scaled marks, y. **[3]**

Cambridge International AS & A Level Mathematics 9709 Paper 73 Q4 June 2015

7 In a survey a random sample of 150 households in Nantville were asked to fill in a questionnaire about household budgeting.

i The results showed that 33 households owned more than one car. Find an approximate 99% confidence interval for the proportion of all households in Nantville with more than one car. **[4]**

ii The results also included the weekly expenditure on food, x dollars, of the households. These were summarised as follows:

$$n = 150, \quad \Sigma x = 19\,035, \quad \text{and} \quad \Sigma x^2 = 4\,054\,716$$

Find unbiased estimates of the mean and variance of the weekly expenditure on food of all households in Nantville. **[3]**

iii The government has a list of all the households in Nantville numbered from 1 to 9526. Describe briefly how to use random numbers to select a sample of 150 households from this list. **[3]**

Cambridge International AS & A Level Mathematics 9709 Paper 72 Q4 November 2014

8 The number of hours that Mrs Hughes spends on her business in a week is normally distributed with mean μ and standard deviation 4.8. In the past the value of μ has been 49.5.

i Assuming that μ is still equal to 49.5, find the probability that in a random sample of 40 weeks the mean time spent on her business in a week is more than 50.3 hours. **[4]**

Following a change in her arrangements, Mrs Hughes wishes to test whether μ has decreased. She chooses a random sample of 40 weeks and notes that the total number of hours she spent on her business during these weeks is 1920.

Copyright Material - Review Only - Not for Redistribution

ii **a** Explain why a one-tail test is appropriate. **[1]**

b Carry out the test at the 6% significance level. **[4]**

c Explain whether it was necessary to use the Central Limit theorem in part **ii b**. **[1]**

Cambridge International AS & A Level Mathematics 9709 Paper 72 Q5 November 2014

PS **9** Following a change in flight schedules, an airline pilot wished to test whether the mean distance that he flies in a week has changed. He noted the distances, x km, that he flew in 50 randomly chosen weeks and summarised the results as follows.

$$n = 50, \ \Sigma x = 143\ 300, \text{ and } \Sigma x^2 = 410\ 900\ 000$$

i Calculate unbiased estimates of the population mean and variance. **[3]**

ii In the past, the mean distance that he flew in a week was 2850 km. Test, at the 5% significance level, whether the mean distance has changed. **[5]**

Cambridge International AS & A Level Mathematics 9709 Paper 71 Q3 November 2013

10 Each of a random sample of 15 students was asked how long they spent revising for an exam.

50 70 80 60 65 110 10 70 75 60 65 45 50 70 50

Assume that the times for all students are normally distributed with mean μ minutes and standard deviation 12 minutes.

i Calculate a 92% confidence interval for μ. **[4]**

ii Explain what is meant by a 92% confidence interval for μ. **[1]**

iii Explain what is meant by saying that a sample is 'random'. **[1]**

Cambridge International AS & A Level Mathematics 9709 Paper 73 Q3 June 2013

11 In the past the weekly profit at a store had mean \$34 600 and standard deviation \$4500. Following a change of ownership, the mean weekly profit for 90 randomly chosen weeks was \$35 400.

i Stating a necessary assumption, test at the 5% significance level whether the mean weekly profit has increased. **[6]**

ii State, with a reason, whether it was necessary to use the Central Limit Theorem in part **i**. **[2]**

The mean weekly profit for another random sample of 90 weeks is found and the same test is carried out at the 5% significance level.

iii State the probability of a Type I error. **[1]**

iv Given that the population mean weekly profit is now \$36 500, calculate the probability of a Type II error. **[5]**

Cambridge International AS & A Level Mathematics 9709 Paper 73 Q7 June 2013

M **12** In order to obtain a random sample of people who live in her town, Jane chooses people at random from the telephone directory for her town.

i Give a reason why Jane's method will not give a random sample of people who live in the town. **[1]**

Jane now uses a valid method to choose a random sample of 200 people from her town and finds that 38 live in apartments.

Copyright Material - Review Only - Not for Redistribution

ii Calculate an approximate 99% confidence interval for the proportion of all people in Jane's town who live in apartments. **[4]**

iii Jane uses the same sample to give a confidence interval of width 0.1 for this proportion. This interval is an x% confidence interval. Find the value of x. **[4]**

Cambridge International AS & A Level Mathematics 9709 Paper 72 Q6 November 2012

PS **13** The volumes of juice in bottles of Apricola are normally distributed. In a random sample of 8 bottles, the volumes of juice, in millilitres, were found to be as follows.

332 334 330 328 331 332 329 333

i Find unbiased estimates of the population mean and variance. **[3]**

A random sample of 50 bottles of Apricola gave unbiased estimates of 331 millilitres and 4.20 millilitres^2 for the population mean and variance respectively.

ii Use this sample of size 50 to calculate a 98% confidence interval for the population mean. **[3]**

iii The manufacturer claims that the mean volume of juice in all bottles is 333 millilitres. State, with a reason, whether your answer to part **ii** supports this claim. **[1]**

Cambridge International AS & A Level Mathematics 9709 Paper 71 Q4 November 2011

PS **14** Metal bolts are produced in large numbers and have lengths which are normally distributed with mean 2.62 cm and standard deviation 0.30 cm.

i Find the probability that a random sample of 45 bolts will have a mean length of more than 2.55 cm. **[3]**

ii The machine making these bolts is given an annual service. This may change the mean length of bolts produced but does not change the standard deviation. To test whether the mean has changed, a random sample of 30 bolts is taken and their lengths noted. The sample mean length is m cm. Find the set of values of m which result in rejection at the 10% significance level of the hypothesis that no change in the mean length has occurred. **[4]**

Cambridge International AS & A Level Mathematics 9709 Paper 71 Q3 June 2010

15 There are 18 people in Millie's class. To choose a person at random she numbers the people in the class from 1 to 18 and presses the random number button on her calculator to obtain a 3-digit decimal. Millie then multiplies the first digit in this decimal by two and chooses the person corresponding to this new number. Decimals in which the first digit is zero are ignored.

i Give a reason why this is not a satisfactory method of choosing a person. **[1]**

Millie obtained a random sample of 5 people of her own age by a satisfactory sampling method and found that their heights in metres were 1.66, 1.68, 1.54, 1.65 and 1.57. Heights are known to be normally distributed with variance 0.0052 m^2.

ii Find a 98% confidence interval for the mean height of people of Millie's age. **[3]**

Cambridge International AS & A Level Mathematics 9709 Paper 72 Q1 November 2009

Copyright Material - Review Only - Not for Redistribution

16 When Sunil travels from his home in England to visit his relatives in India, his journey is in four stages. The times, in hours, for the stages have independent normal distributions as follows:

Bus from home to the airport: N(3.75, 1.45)

Waiting in the airport: N(3.1, 0.785)

Flight from England to India: N(11, 1.3)

Car in India to relatives: N(3.2, 0.81)

i Find the probability that the flight time is shorter than the total time for the other three stages. **[6]**

ii Find the probability that, for 6 journeys to India, the mean time waiting in the airport is less than 4 hours. **[3]**

Cambridge International AS & A Level Mathematics 9709 Paper 71 Q6 June 2009

PS **17** The times taken for the pupils in Ming's year group to do their English homework have a normal distribution with standard deviation 15.7 minutes. A teacher estimates that the mean time is 42 minutes. The times taken by a random sample of 3 students from the year group were 27, 35 and 43 minutes. Carry out a hypothesis test at the 10% significance level to determine whether the teacher's estimate for the mean should be accepted, stating the null and alternative hypotheses. **[5]**

Cambridge International AS & A Level Mathematics 9709 Paper 7 Q2 November 2008

M **18** Diameters of golf balls are known to be normally distributed with mean μ cm and standard deviation σ cm. A random sample of 130 golf balls was taken and the diameters, x cm, were measured. The results are summarised by $\sum x = 555.1$ and $\sum x^2 = 2371.30$.

i Calculate unbiased estimates of μ and σ^2. **[3]**

ii Calculate a 97% confidence interval for μ. **[3]**

iii 300 random samples of 130 balls are taken and a 97% confidence interval is calculated for each sample. How many of these intervals would you expect **not** to contain μ? **[1]**

Cambridge International AS & A Level Mathematics 9709 Paper 7 Q4 November 2008

PS **19** The time in hours taken for clothes to dry can be modelled by the continuous random variable with probability density function given by:

$$f(t) = \begin{cases} k\sqrt{t} & 1 \leqslant t \leqslant 4, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

i Show that $k = \frac{3}{14}$. **[3]**

ii Find the mean time taken for clothes to dry. **[4]**

iii Find the median time taken for clothes to dry. **[3]**

iv Find the probability that the time taken for clothes to dry is between the mean time and the median time. **[2]**

Cambridge International AS & A Level Mathematics 9709 Paper 7 Q7 November 2008

Copyright Material - Review Only - Not for Redistribution

20 A magazine conducted a survey about the sleeping time of adults. A random sample of 12 adults was chosen from the adults travelling to work on a train.

i Give a reason why this is an unsatisfactory sample for the purposes of the survey. **[1]**

ii State a population for which this sample would be satisfactory. **[1]**

A satisfactory sample of 12 adults gave numbers of hours of sleep as shown below.

4.6 6.8 5.2 6.2 5.7 7.1 6.3 5.6 7.0 5.8 6.5 7.2

iii Calculate unbiased estimates of the mean and variance of the sleeping times of adults. **[3]**

Cambridge International AS & A Level Mathematics 9709 Paper 7 Q1 June 2008

Copyright Material - Review Only - Not for Redistribution

PRACTICE EXAM-STYLE PAPER

Time allowed is 1 hour and 15 minutes (50 marks).

1 Telephone calls to a company are answered within 3 minutes; if they are not answered within 3 minutes, they are disconnected. The length of time before calls are answered, T minutes, is modelled using the probability density function:

$$f(t) = \begin{cases} kt(10 - 3t) & 0.5 \leqslant t \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant.

a Sketch the graph of $y = f(t)$ and show that $k = \frac{8}{135}$. **[5]**

b **i** Find the probability that a randomly chosen caller waits longer than 2 minutes for their call to be answered. **[2]**

ii For ten randomly chosen callers, find the probability that no more than one caller waits for longer than 2 minutes for their call to be answered. You may assume that the waiting times are independent. **[3]**

c Find the mean waiting time. **[3]**

d Explain whether or not the probability density function for T is realistic. **[1]**

2 The number, X, of breakdowns per week of escalators in a large shopping mall has a Poisson distribution with mean 0.9.

a Find the probability that there will be:

i exactly three breakdowns during one week

ii exactly three breakdowns each week for three successive weeks

iii at most two breakdowns in a particular week. **[6]**

b Find the probability that during a particular four-week period there will be no breakdowns. **[2]**

c After the maintenance contract for the escalators is given to a new company, it is found there is one breakdown in a four-week period. Perform a 10% significance test to determine whether the average number of breakdowns has decreased. **[5]**

3 A fair coin is flipped 12 times. Heads appears two times. In a significance test for bias towards tails, the calculated probability results in a Type I error. Find the minimum level of significance of the test, giving your answer as an integer. **[4]**

4 The maker of a certain brand of chocolate sells boxes of individually wrapped chocolates. Weights of chocolates follow a normal distribution with mean 17 g and standard deviation 2 g. Weights of wrapping have an independent normal distribution with mean 4.5 g and standard deviation 0.3 g. Eight of these individually wrapped chocolates are placed together in a box. The weight of the box is 50 g. Find the probability that a randomly chosen box of chocolates weighs less than 230 g. **[5]**

5 Records from a physiotherapy clinic show that in one year the length of time for which patients are treated during a visit is normally distributed with mean 32.2 minutes and standard deviation 3.9.

a Calculate an interval within which 95% of the times will lie. **[3]**

b The following year a random sample of ten physiotherapy patients' visits have the following times (in minutes):

28.2 27.6 32.4 33.1 35.0 26.9 28.5 25.4 32.8 31.7

Copyright Material - Review Only - Not for Redistribution

Assuming the times are normally distributed and the standard deviation remains unchanged from the previous year, use these times to calculate a 90% confidence interval for the population mean. **[4]**

6 A manufacturer claims the average lifetime of a particular battery it produces for smoke alarms is 76 weeks with a standard deviation of 4 weeks. To test this claim a random sample of 40 batteries is tested and found to have a mean lifetime of 74.6 weeks.

Stating clearly any assumptions you make, and using a 2% level of significance, test the claim made by the manufacturer. **[7]**

Review Copy - Cambridge University Press - Review Copy

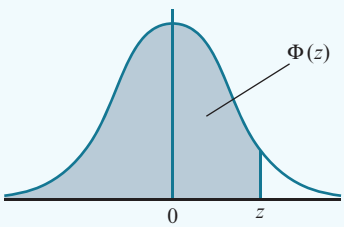
Copyright Material - Review Only - Not for Redistribution

THE STANDARD NORMAL DISTRIBUTION FUNCTION

If Z is normally distributed with mean 0 and variance 1, the table gives the value of $\Phi(z)$ for each value of z, where

$$\Phi(z) = \mathrm{P}(Z \leqslant z).$$

Use $\Phi(-z) = 1 - \Phi(z)$ for negative values of z.



z	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
											ADD								
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	4	8	12	16	20	24	28	32	36
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	4	8	12	16	20	24	28	32	36
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	4	8	12	15	19	23	27	31	35
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	4	7	11	15	19	22	26	30	34
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	4	7	11	14	18	22	25	29	32
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	3	7	10	14	17	20	24	27	31
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	3	7	10	13	16	19	23	26	29
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	3	6	9	12	15	18	21	24	27
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	3	5	8	11	14	16	19	22	25
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	3	5	8	10	13	15	18	20	23
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	2	5	7	9	12	14	16	19	21
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	2	4	6	8	10	12	14	16	18
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	2	4	6	7	9	11	13	15	17
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	2	3	5	6	8	10	11	13	14
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	1	3	4	6	7	8	10	11	13
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	1	2	4	5	6	7	8	10	11
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	1	2	3	4	5	6	7	8	9
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	1	2	3	4	4	5	6	7	8
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	1	1	2	3	4	4	5	6	6
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	1	1	2	2	3	4	4	5	5
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	0	1	1	2	2	3	3	4	4
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	0	1	1	2	2	2	3	3	4
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	0	1	1	1	2	2	2	3	3
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	0	1	1	1	1	2	2	2	2
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	0	0	1	1	1	1	1	2	2
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	0	0	0	1	1	1	1	1	1
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	0	0	0	0	1	1	1	1	1
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	0	0	0	0	0	1	1	1	1
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	0	0	0	0	0	0	0	1	1
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	0	0	0	0	0	0	0	0	0

Copyright Material - Review Only - Not for Redistribution

Critical values for the normal distribution

The table gives the value of z such that $P(Z \leqslant z) = p$, where $Z \sim N(0, 1)$.

p	0.75	0.90	0.95	0.975	0.99	0.995	0.9975	0.999	0.9995
z	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Review Copy - Cambridge University Press - Review Copy

Copyright Material - Review Only - Not for Redistribution

Answers

1 Hypothesis testing

Prerequisite knowledge

1 **a** 0.149 **b** 0.813 **c** 0.005

2 **a** 0.0393 to 3 s.f. **b** 0.0802 to 3 s.f. **c** 0.381

3 **a** 0.692 **b** 0.960 to 3 s.f. **c** 0.440

4 **a** 0.122 **b** 0.798

5 $X \sim N(32, 19.2)$

a 0.81948 **b** 0.634

6 $X \sim N(66, 29.7)$

a 0.205 **b** 0.26

Exercise 1A

1 **a** Let X be the number of nature programmes, $X \sim B(20, 0.25)$; $H_0: p = 0.25$; $H_1: p < 0.25$

b $P(X \leqslant 2) = 0.0913$. As this is less than 10%, reject H_0. There are fewer nature programmes than claimed.

c 0.0913 is greater than 5%, therefore accept H_0. There is insufficient evidence to reject the claim that 25% of programmes are nature programmes.

2 Let X be the number of footballers who cannot explain the offside rule, $X \sim B(15, 0.4)$; $H_0: p = 0.4$; $H_1: p < 0.4$

$P(X \leqslant 2) = 0.0271$. As this is less than 5%, reject H_0. There is some evidence to support the claim that the number of professional footballers who do not understand the offside rule is less than 40%.

3 **a** Let X be the number of sheep deficient in a given mineral, $X \sim B(80, 0.3)$; $H_0: p = 0.3$; $H_1: p < 0.3$

$P(X \leqslant 19) = 0.136$. As this is greater than 10%, accept H_0. There is insufficient evidence to support the claim that the percentage of sheep deficient in a particular mineral has decreased.

b 18

4 **a** $P(X \geqslant 7) = 0.0898$. As this is greater than 5%, accept H_0. There is insufficient evidence to support the claim that the coin is biased towards tails.

b $Y \sim B(9, 0.5)$; $H_0: p = 0.5$; $H_1: p < 0.5$

$P(X \leqslant 2) = 0.0898$

This is >5%, therefore accept H_0.

There is insufficient evidence to reject the claim.

c $X \sim N(90, 45)$; $H_0: \mu = 90$; $H_1: \mu > 90$

$P(X \geqslant 102) = 0.0432$. As this is less than 10%, reject H_0. There is some evidence to support the claim that the coin is biased towards tails.

5 $X \sim B(120, 0.25)$; $H_0: p = 0.25$; $H_1: p > 0.25$ or $X \sim N(30, 22.5)$; $H_0: \mu = 30$; $H_1: \mu > 30$

$P(X \geqslant 40) = 0.0226$. As this is greater than 2%, accept H_0. There is insufficient evidence to support the claim that the spinner is biased towards blue.

6 $X \sim B(17, 0.6)$; $H_0: p = 0.6$; $H_1: p > 0.6$

Test statistic is $P(X \geqslant 15) = 0.0123$.

Significance level is 2%.

Exercise 1B

1 **a** $X \sim B(20, 0.25)$; $H_0: p = 0.25$; $H_1: p \neq 0.25$

b $P(X \leqslant 2) = 0.0913$. As this is greater than 5%, accept H_0. There is insufficient evidence to reject the claim that 25% of programmes are nature programmes.

2 $X \sim B(15, 0.4)$; $H_0: p = 0.4$; $H_1: p \neq 0.4$

$P(X \leqslant 2) = 0.0271$. As this is greater than 2.5%, accept H_0. There is insufficient evidence to reject the claim that 40% of professional footballers do not understand the offside rule.

3 **a** $X \sim B(16, 0.8)$; $H_0: p = 0.8$; $H_1: p \neq 0.8$

$P(X \geqslant 15) = 0.141$. As this is greater than 5%, accept H_0. There is insufficient evidence to reject the claim that 80% of Americans believe in horoscopes.

Critical value = 16

b $X \sim B(160, 0.8)$; $H_0: p = 0.8$; $H_1: p \neq 0.8$ or $X \sim N(128, 25.6)$; $H_0: \mu = 128$; $H_1: \mu \neq 128$

$P(X \leqslant 116) = 0.0115$. As this is less than 2.5%, reject H_0. There is sufficient evidence to reject the claim that 80% of Americans believe in horoscopes.

4 **a** $X \sim B(180, 0.12)$; $H_0: p = 0.12$; $H_1: p \neq 0.12$ or $X \sim N(21.6, 19.008)$; $H_0: \mu = 21.6$; $H_1: \mu \neq 21.6$

$P(X \leqslant 13) = 0.0316$

Because $P(X \leqslant 13)$ is less than 0.04, and $P(X \leqslant 14)$ is more than 0.04, the significance level is 8% for the two tailed test (4% in each tail). By definition, if 13 is the critical value, we have to reject H_0, and we must have 0.0318 less than the significance level divided by two.

There is evidence to suspect the percentage of people who can spell onomatopoeia is incorrect.

Copyright Material - Review Only - Not for Redistribution

b 0.0318 < 8%, therefore reject H_0. There is some evidence to suggest that a lower percentage of people can spell onomatopoeia.

5 **a** $X \sim B(20, 0.3)$; H_0: $p = 0.3$; H_1: $p \neq 0.3$

$P(X \leqslant 3) = 0.107$. As this is greater than 5 %, accept H_0. There is insufficient evidence to reject the claim that 30% are red.

Critical value = 2

b H_0: $p = 0.3$; H_1: $p < 0.3$

No, Ginny's conclusion will be the same.

6 $X \sim B(600, 0.01)$; H_0: $p = 0.01$; H_1: $p \neq 0.01$ or $X \sim N(6, 5.94)$; H_0: $\mu = 6$; H_1: $\mu \neq 6$

$P(X \leqslant 11) = 0.988$. $1 - 0.988 < 2\%$, reject H_0. There is sufficient evidence to accept the claim.

Exercise 1C

1 **a** $X \sim B(12, 0.4)$

b H_0: $p = 0.4$; H_1: $p < 0.4$

c $P(X \leqslant a) < 5\%$

$P(X \leqslant 1) = 0.0196 < 5\%$, whereas $P(X \leqslant 2) = 0.0835 > 5\%$.

So the rejection region is $X \leqslant 1$

d So you reject the archer's claim if $X = 1$.

2 **a** $X \sim B(16, 0.8)$

b H_0: $p = 0.8$; H_1: $p > 0.8$

c $P(X >= a) < 10\%$

d $P(X \geqslant 15) = 0.1407 > 10\%$, whereas $P(X \geqslant 16) = 0.0281 < 10\%$. So the rejection region is $X \geqslant 16$.

So the probability of a Type I error is 0.0281.

3 $X \sim B(300, 0.2)$; H_0: $p = 0.2$; H_1: $p < 0.2$

$X \sim N(60, 48)$; H_0: $\mu = 60$; H_1: $\mu < 60$

$P(X \leqslant a - 0.5) < 5\%$

$$\frac{a - 0.5 - 60}{\sqrt{48}} = -1.645$$

$$a = 49.1$$

49 teenagers

4 **a** H_0: $\mu = 1.8$; H_1: $\mu > 1.8$

b $$1 - \Phi\left(\frac{2.25 - 1.8}{0.32}\right) = 1 - \Phi(1.40625) = 1 - 0.9202 = 0.0798$$

c $$\Phi\left(\frac{2.25 - 2.4}{0.32}\right) = \Phi(-0.46875) = 1 - 0.6804 = 0.320$$

5 **a** Die is considered to be biased when it is in fact fair.

$X \sim B(20, 0.125)$

$P(X \geqslant 4) = 1 - P(X \leqslant 3) = 1 - 0.765 = 0.235$

b Die is considered to be fair when it is in fact biased.

Need to know the probability of rolling a 1 with a biased die.

End-of-chapter review exercise 1

1 **a** $X \sim B(25, 0.16)$; H_0: $p = 0.16$; H_1: $p < 0.16$

$P(X \leqslant 1) = 0.0737 > 5\%$, so accept H_0.

b $X \sim B(150, 0.16)$

$X \sim N(24, 20.16)$; H_0: $\mu = 24$; H_1: $\mu < 24$

$$\Phi\left(\frac{18.5 - 24}{\sqrt{20.16}}\right) = \Phi(-1.225) = 1 - 0.8897 = 0.1103 > 5\%$$

So accept H_0.

There is no evidence to suggest that the new clay is more reliable.

2 $X \sim B(1000, 0.2)$; $X \sim N(200, 160)$; H_0: $\mu = 200$; H_1: $\mu > 200$

$$\Phi\left(\frac{217.5 - 200}{\sqrt{160}}\right) = \Phi(1.383) = 0.9167 < 95\%$$

So accept H_0.

There is no evidence to suggest the fens will flood more often.

3 **a** $X \sim N(800, 50^2)$; H_0: $\mu = 800$; H_1: $\mu \neq 800$

$$\Phi\left(\frac{720 - 800}{50}\right) = 0.0548$$

The light bulb lasting 720 hours is fewer hours than claimed, but the null hypothesis is accepted and if you use a light bulb that lasts this long you could not statistically say the manufacturer's claim is unjustified.

$$1 - \Phi\left(\frac{920 - 800}{50}\right) = 0.00820$$

The light bulb lasting 920 hours statistically lasts longer than the manufacturer's claims; you would reject H_0 and accept H_1. In this case, however, your light bulb is longer-lasting than expected, so you may not want to challenge the manufacturer's claim.

b 702 hours

4 $X \sim B(40, 0.08)$

$P(X \leqslant 0) = 0.0356 < 10\%$

$P(X \leqslant 1) = 0.1594 > 10\%$

Critical region $X \leqslant 0$

Critical value $X = 0$

Copyright Material - Review Only - Not for Redistribution

5 **a** $X \sim B(850, 0.012)$; $X \sim N(10.2, 10.0776)$;
$H_0: \mu = 10.2$; $H_1: \mu > 10.2$
$1 - \Phi\left(\frac{14.5 - 10.2}{\sqrt{10.0776}}\right) = 1 - \Phi(1.355) = 1 - 0.9123 > 2\%$
So accept H_0.
There is no evidence to suggest there is a greater proportion of red-haired people in Scotland.

b $1 - \Phi\left(\frac{x - 0.5 - 10.2}{\sqrt{10.0776}}\right) < 0.02$
$\frac{x - 0.5 - 10.2}{\sqrt{10.0776}} > 2.054$
Critical value = 18

6 **i** $H_0: p = 0.3$; $H_1: p < 0.3$
ii Probability of Type I error = 0.0355
iii There is insufficient evidence to suggest the proportion of packets containing gifts is less than 30%.

7 **i** $H_0: p = \frac{1}{8}$; $H_1: p > \frac{1}{8}$
ii 0.120 **iii** 12%

8 **i** There is insufficient evidence to support the claim.
ii Normal; $\mu = 200$; $\sigma^2 = 160$ or $\sigma = \sqrt{160}$
iii Type II error = accepting the hypothesis that the machine produces the correct number of 5s when it does not.

9 **i** Fewer than 2 samples contain a free gift.
ii 0.0243
iii Two is not in the critical region, so there is insufficient evidence to reject the claim.

10 There is insufficient evidence to support the claim.

11 **i** $H_0: p = \frac{1}{6}$; $H_1: p > \frac{1}{6}$
ii 0.0697
iii A Type II error occurs if the die is biased towards six, but less than 4 sixes are thrown, so there is no evidence of bias.
iv 0.172 or $\frac{11}{64}$

12 There is insufficient evidence to support the claim that women in Jakarta have smaller fingers.

2 The Poisson distribution

Prerequisite knowledge

1 **a** 0.000786 **b** 0.376

2 **a** 0.999 **b** 0.215

3 **a** 0.135 **b** 0.0273 **c** 0.000123

4 $H_0: \mu = 42$; $H_1: \mu \neq 42$; $\Phi\left(\frac{45 - 42}{\sqrt{8}}\right)$

Exercise 2A

1 **a** 0.156 **b** 0.238 **c** 0.982

2 **a** 0.149 **b** 0.970

3 **a** 0.0471 **b** 0.809 **c** 0.191

4 **a** 0.125
b Occurrence of potholes may be affected by factors such as materials used, quality of work, weather, etc.

5 **a** **i** Calls occur singly, independently and at random.
ii Unlikely as more than one person may call to report a large fire; hence, calls may not always be independent.
b 0.0153

6 **a** **i** 0.134 **ii** 0.928
b Choose some bars of the same hazelnut chocolate at random; count the number of whole hazelnuts; draw a frequency table and calculate mean and variance.

7 **a** Mean 1.625 or 1.63; variance 1.61; mean and variance are similar.
b 13, 20, 17, 9, 4, 1, 0

Exercise 2B

1 **a** 0.113 **b** 0.000499

2 0.0907

3 Flaws occur independently and at random.
a 0.000335 **b** 0.191

4 **a** 0.958
b Possible accident or road closure, or the random 20 minutes happened to be in the middle of the night.
c Traffic flow different, cars not independent.

5 **a** 0.0183 **b** 0.938

6 **a** 0.406 **b** 0.0916

7 **a** Orders are placed independently and at random.
b **i** 0.287 **ii** 0.358

8 **a** Emails are received independently and at random.
b A faulty product may generate lots of emails.
c People unlikely to complain at night.
d **i** 0.129 **ii** 0.000053

Exercise 2C

1 **a** 0.647 **b** 0.353

2 **a** 0.964 **b** 0.430

Copyright Material - Review Only - Not for Redistribution

3 0.00449

4 0.937

5 $X \sim \text{Po}(1.2)$; 0.966

6 **a** 0.0656

b Accept 13 800 or 13 816.

7 **a** n large, p small

b 0.294 **c** 0.246

Exercise 2D

1 **a** 0.876 **b** 0.409 **c** 0.0552

2 **a** 0.144 **b** 0.167

3 0.186

4 **a** 0.119

b No; the reason for delay (e.g. weather/accident) can affect delays on subsequent days.

5 0.964

6 50

7 51

8 **a** 0.0668

b $\Phi\left(\frac{38.5-25}{\sqrt{25}}\right)-\left(1-\Phi\left(\frac{11.5-25}{\sqrt{25}}\right)\right)$

$= 0.9965-(1-0.9965) = 0.993$

Exercise 2E

1 $H_0: \lambda = 62$ or $\mu = 62$

$H_1: \lambda > 62$ or $\mu > 62$

Probability = 0.170

Accept H_0.

2 $H_0: \lambda = 7$; $H_1: \lambda < 7$

Probability = 0.173

Accept H_0.

3 $H_0: \lambda = 39$ or $\mu = 39$; $H_1: \lambda < 39$ or $\mu < 39$

Probability = 0.0464

Reject H_0.

4 **a** $H_0: \lambda = 64$ or $\mu = 64$; $H_1: \lambda \neq 64$ or $\mu \neq 64$

Probability = 0.0349

Accept H_0.

b 47 misprints

End-of-chapter review exercise 2

1 **a** 0.195 **b** 0.762

c 0.122 **d** 0.809

2 **a** Mean = 3.425, variance = 3.21. These are fairly close, using a Poisson distribution is appropriate.

Could also use the mean to calculate expected frequencies for comparison.

b 0.433

3 **a** 0.0796 **b** 0.0687 **c** 0.385

4 $H_0: \lambda = 3$; $H_1: \lambda > 3$

Probability $(X \geqslant 5) = 0.185 > 5\%$

This is not in the critical region, so accept H_0.

5 0.353

6 **i** 0.135

ii **a** 0.744 **b** 0.870

iii 6

7 **i** Constant mean rate **ii** 0.0922

iii 0.821 **iv** 0.0723

8 **i** 0.469 **ii** 2.15 minutes

9 **i** Customers arrive independently.

ii 0.161 **iii** 0.570 **iv** 0.869

10 **i** Constant average rate of goals scored; goals scored independently.

ii 0.259 **iii** 0.164

11 **i** $P(X \geq 5) = 0.0959 > 0.05$ There is insufficient evidence to suggest mean sales have increased.

ii 0.0357

iii Mean sales = 0.8 per week but 6 or more sales in 3 weeks leads to rejecting mean = 0.8.

iv The value of the new mean.

12 **i** Type I error shows the number of white blood cells has decreased when in fact it has not decreased. Probability = 0.0342

ii There is insufficient evidence to show a decrease in the number of white blood cells.

iii 0.915

13 **i** 0.143 **ii** 0.118 **iii** 0.0316

3 Linear combinations of random variables

Prerequisite knowledge

1 $E(X) = 2\frac{1}{6}$; $\text{Var}(X) = \frac{29}{36}$

2 **a** 0.673 **b** 0.673

3 **a** 0.119 **b** 0.0289

Exercise 3A

1 **a** 74 **b** 150

2 **a** 15 **b** 19.2

Copyright Material - Review Only - Not for Redistribution

3 **a** **i** 9.2 **ii** 7.2

b $a = 5, b = 20$ or $a = -5, b = 44$

4 **a** **i** 19 **ii** 20.8

b $a = 10, b = 30$ or $a = -10, b = -70$

5 **a** 8.5 **b** 15.25

6 **a** 11 **b** $46\frac{2}{3}$

7 **a** 25.8 **b** 310

8 $a = \frac{1}{\sigma}, b = \frac{\mu}{\sigma}$ or $a = \frac{-1}{\sigma}, b = \frac{-\mu}{\sigma}$

9 Mean = 76.1; variance = 6.48

Exercise 3B

1 **a** 7.2 **b** 19.68

2 **a** 18 **b** 18.3

3 **a** 39; 6.6 **b** −8; 3.5

4 **a** 22.5 **b** 14.6

5 **a** 20 **b** 4.3 **c** 7 **d** 16.2

6 **a** 3; 1.8 **b** 6; 7.2 **c** 6 **d** 3.6

7 Mean = 11; variance = $51\frac{1}{6}$; sd = 7.15

Exercise 3C

1 **a** $A + 2B \sim N(66, 104)$

b $2B - A \sim N(46, 104)$

c $A + B + C \sim N(53, 38)$

2 **a** $2X - Y \sim N(23, 5^2)$ **b** 0.841

3 **a** 0.421 **b** 0.788

4 0.401

5 7020; 76.8

6 0.0455

7 0.607

8 **a** N(1.6, 0.0008) **b** N(1.6, 0.0128)

Exercise 3D

1 **a** 2.1; 23.1 **b** 10.5; 10.5

c 23.1; 60.9 **d** **b** has a Poisson distribution.

2 **a** 12 **b** 0.000442

3 0.00470

4 **a** 0.128 **b** 0.0111

5 $T \sim Po(3)$; 0.448

6 **a** 0.0361 **b** 0.594

7 **a** $T \sim Po(3.7)$

b **i** 0.193 **ii** 0.285

8 0.175

9 Assume items are received at random.

a 0.215 **b** 0.228

10 0.704. It is necessary to assume that goals scored by each team are independent. Not reasonable since the teams are opponents.

End-of-chapter review exercise 3

1 63.7; 3.6

2 0.747

3 0.149

4 0.542

5 0.721

6 0.0356

7 **i** 0.889

ii 0.176

8 **i** 9.3 **ii** 27.9

iii Mean = −2.0; variance = 37.2

9 **i** Mean = 926 cents; sd = 105 cents

ii Mean = 5556 cents; sd = 257 cents

10 0.9405

11 **i** 0.823

ii Mean = 376 miles; sd = 39.0 miles

12 **i** $a = -6.32; b = 2$

ii **a** 1.87 **b** 0.699

Cross-topic review exercise 1

1 **a** 0.951

b 0.00709

c $0.967 < 0.98$

There is insufficient evidence to reject the claim.

2 0.336

3 $\lambda + 2.326\sqrt{\lambda} - 55.5 = 0$

$\lambda = 40.7$

4 **i** 0.216 **ii** 0.973 **iii** 0.221

5 **i** 0.186 **ii** 0.257 **iii** 0.191

iv The officer will reject H_0.

6 **i** 0.0254 **ii** 0.983

7 **i** 0.256

ii 0.117

iii Type I: Luigi's belief is rejected, though it may be true.

Copyright Material - Review Only - Not for Redistribution

8 **i** 0.259 **ii** $e^{-2} \times \frac{2^r}{r!} = \frac{2}{3}e^{-2}$

$3 \times 2^3 = 4!$

9 **i** 0.294 **ii** 5 **iii** 0.0753

10 **i** 0.191 **ii** 0.112 **iii** 3

11 **i** 0.202 **ii** 0.159

12 **i** 0.0202 **ii** 0.972 **iii** 0.0311

13 **i** 0.938 **ii** 0.993

14 **i** 5 years = 60 months; 1 in 15 ≡ 4 in 60

ii 0 or 1

iii 0.0916

iv 1 lies in the rejection region; there is evidence that the new guitar strings last longer.

4 Continuous random variables

Prerequisite knowledge

1 22

2 0.8

3 10

4 0.785

5 $m = 0.5$

Exercise 4A

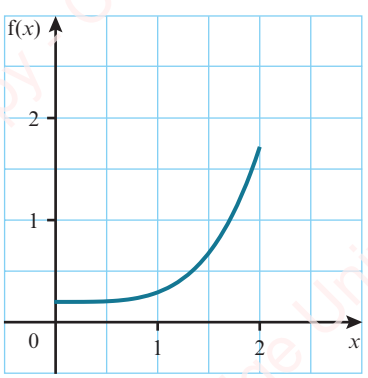
1 **a** $\int_0^2 kx^4 + \frac{1}{5}\,dx = 1$

$\left[\frac{kx^5}{5} + \frac{x}{5}\right]_0^2 = 1$

$\left(\frac{32k}{5} + \frac{2}{5}\right) = 1$

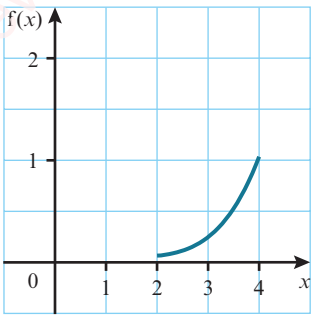
$k = \frac{3}{32}$

b



c $\frac{7}{32}$

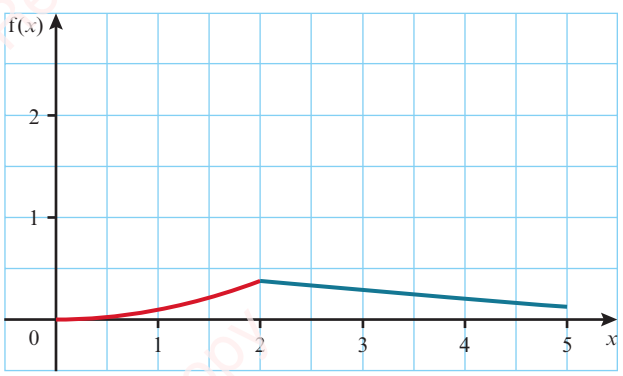
2 **a** $\int_2^4 \frac{x^3}{60}\,dx = \left[\frac{x^4}{240}\right]_2^4 = \frac{256}{240} - \frac{16}{240} = 1$



b $\frac{35}{48}$ **c** 0

3 **a** $\int_0^2 \frac{3x^2}{32}\,dx + \int_2^5 \frac{1}{24}(13 - 2x)\,dx$

$= \left[\frac{x^3}{32}\right]_0^2 + \left[\frac{1}{24}(13x - x^2)\right]_2^5 = 1$

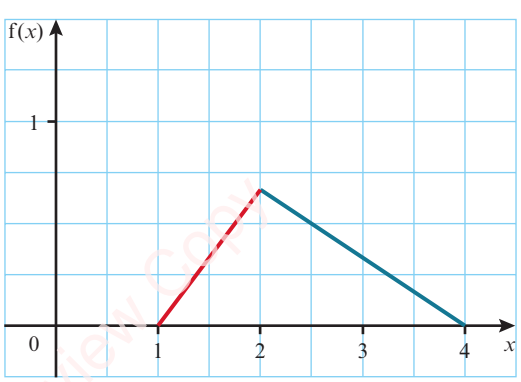


b **i** $\frac{1}{6}$

ii $\frac{53}{96}$ or 0.552

4 **a** $\int_1^2 \frac{2(x-1)}{3}\,dx + \int_2^4 \frac{4-x}{3}\,dx$

$= \left[\frac{x^2 - 2x}{3}\right]_1^2 + \left[\frac{4x}{3} - \frac{x^2}{6}\right]_2^4 = 1$



b $\frac{5}{6}$

Copyright Material - Review Only - Not for Redistribution

c $k = 2$. $\frac{1}{3}$ corresponds to the area of the triangle rectangle defined by the red line and $1 < x < 2$.

5 1.25

6 $\frac{11}{40}$

7 0.0183

Exercise 4B

1 **a** $\int_2^5 k(x-2)\,dx + \int_5^7 k(7-x)\,dx = 1$

b $\frac{4}{13}$ **c** $\frac{11}{13}$

d $2 + \frac{\sqrt{26}}{2}$ **e** $2 + \frac{\sqrt{65}}{5}$

2 **a** $\frac{16}{25}$

b Solve $m^2 - 10m + 12.5 = 0$; other root outside range.

c $\frac{5}{2}\left(\sqrt{3} - 1\right)$

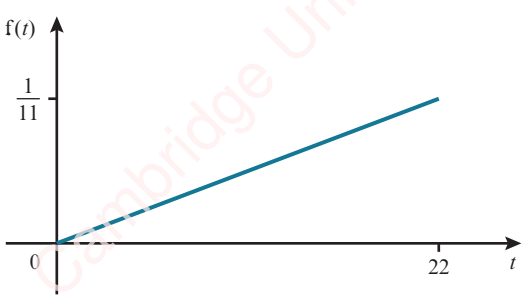
3 **a** Sketch or $\int_{10}^{30} \frac{1}{k}(40-t)\,dt = 1$

b $\frac{11}{32}$ or 0.34375

c $40 - 10\sqrt{5}$

d Likely to talk for shorter or longer periods.

4 **a**



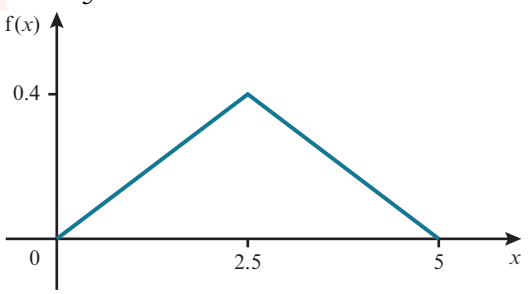
b 0.948 or $\frac{459}{484}$

c Either integrate with limits from 0 to 20 (gives probability 0.826 or $\frac{100}{141} > 80\%$) or attempt to find upper value with integral = 0.8, upper value 19.7 (3 significant figures).

5 **a** Graphically or $\int_0^{2.5} kx\,dx + \int_{2.5}^5 k(5-x)\,dx = 1$

b 0.6 or $\frac{3}{5}$

c



i Peak at 2.5

ii Mid point of symmetrical graph, 2.5

6 **a** 0.812 **b** 800

Exercise 4C

1 Mean = 2.5; variance = 1.25

2 Mean = 2.81; variance = 1.13

3 **a** $k\left(\left[\frac{x^3}{3}\right]_0^2 + \left[6x - \frac{x^2}{2}\right]_2^6\right) = 1$

$k\left(\frac{8}{3} + 18 - 10\right) = 1$

$k = \frac{3}{32}$ as required

b Mean = 2.88; variance = 1.33

4 **a** $k\left[x^3 - 2x^2 + 2x\right]_3^5 = 1$

$k(85 - 15) = 1$

$k = \frac{1}{70}$ as required

b 4.19 kg **c** \$264

5 **a** 0.4 **b** 4.08 or 4

6 **a** Mean = $\frac{287}{150}$ or 1.913; sd = 0.522

b 0.510

c 600

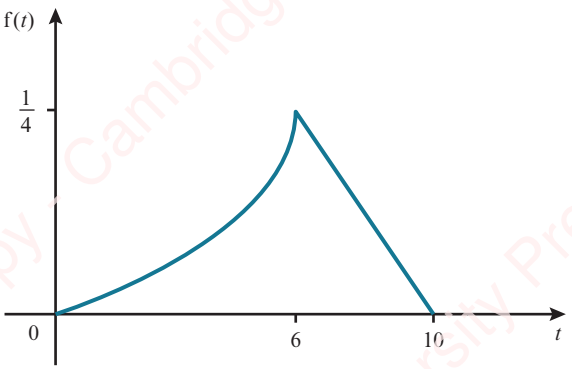
7 **a** 0.741

b Mean = 5; variance = 25

Copyright Material - Review Only - Not for Redistribution

End-of-chapter review exercise 4

1 **a**



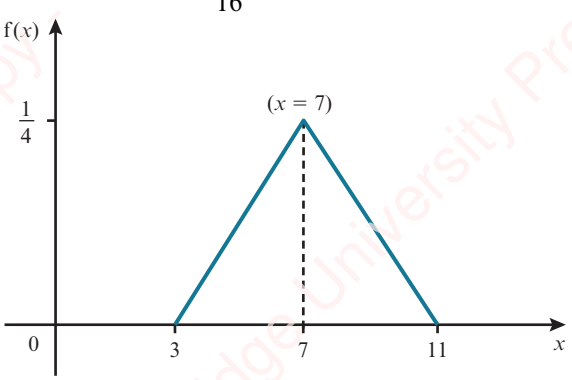
b $\frac{23}{32}$ **c** 5 min 55 s

2 **a** $k\int_3^7 x-3\,\mathrm{d}x + k\int_7^{11} 11-x\,\mathrm{d}x = 1$

$k\left(\left[\frac{x^2}{2}-3x\right]_3^7 + \left[11x-\frac{x^2}{2}\right]_7^{11}\right) = 1$

$16k = 1 \quad k = \frac{1}{16}$ as required

b

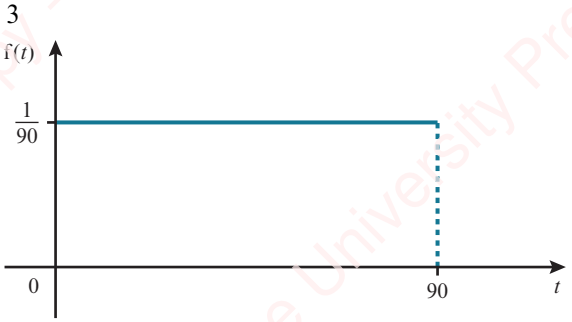


The sketched distribution is symmetrical about $x = 7$.

c $11-2\sqrt{2}$ **d** $\frac{8}{3}$

3 **a** $\frac{1}{3}$

b



The graph is uniform.

c 45 min

d 0.286

e 0.0953

4 **a** $\frac{14}{27}$

b Mean $= \frac{9}{8}$; var $= \frac{171}{320}$

c 0.6603 (Integrate from 0 to $\frac{10}{7}$.)

5 **i** X, Z, Y, W or $\sigma_X, \sigma_Z, \sigma_Y, \sigma_W$

ii **a** Mean $= 0$

$\frac{1}{18}\int_{-3}^{3} x^4\,\mathrm{d}x - 0$

$= \frac{1}{18}\left[\frac{x^5}{5}\right]_{-3}^{3}$

$= 5.4$

$\sqrt{5.4} = 2.32$, as required.

b 0.268

c 0

6 **a** 0.8

b **i** $\int_0^{1.5} k\left(2.25-x^2\right)\mathrm{d}x = 1$

$k\left[2.25x-\frac{x^3}{3}\right]_0^{1.5} = 1$

$k = \frac{4}{9}$, as required.

ii 56 h 15 min

iii Max value of x is 150 hours, whereas daffodils can last for 290 hours.

iv 2.9 to 5 inclusive

7 **a** $\frac{5}{32}$

b **i** $\sqrt{\frac{2}{\pi}}$

ii 0

iii 0.8

8 $2\sqrt{2}$

9 **i** Longest lifetime

ii $\int_1^a \frac{k}{x^2}\,\mathrm{d}x = 1$

$\left[\frac{-k}{x}\right]_1^a = 1$

$k = \frac{a}{a-1}$, as required.

iii 1.53

10 **i** $\int_0^{\frac{2\pi}{3}} k\sin x\,\mathrm{d}x = 1$

$\left[-k\cos x\right]_0^{\frac{2\pi}{3}} = 1$

$k = \frac{2}{3}$, as required.

Copyright Material - Review Only - Not for Redistribution

ii $\frac{2}{3}\int_0^m \sin x\,dx = 0.5$

$\frac{2}{3}\left[-\cos x\right]_0^m = 0.5$

$\frac{2}{3}(-\cos m + 1) = 0.5$

$\cos m = 0.25$

$m = 1.32$

iii 1.28 to 3 s.f.

11 **i** $\int_{-1}^{1} k(1-x)\,dx = 1$

$k\left[x - \frac{x^2}{2}\right]_{-1}^{1} = 1$

$k = \frac{1}{2}$, as required.

ii 0.0625

iii $-\frac{1}{3}$

iv $1 - \sqrt{3}$

12 **a** $\int_3^6 k(6t - t^2)\,dt = 1$

$k\left[3t^2 - \frac{t^3}{3}\right]_3^6 = 1$

$k = \frac{1}{18}$, as required.

b $\frac{33}{8}$

c $\frac{4}{27}$

d Less than 5 min, as answer to part **c** $< \frac{1}{4}$.

5 Sampling

Prerequisite knowledge

1 **a** 0.894 **b** 0.290

2 **a** 0.807 **b** 0.0745

Exercise 5A

1 **a** 713 299 680 324 413 623

b 039 282 237 992 325 799

2 **a** Reads across each row then down the column.

b Reads two values in each line.

c Reads down the column.

3 **a** This method would give only those people who arrive early to work the chance of being selected. It would also exclude anyone working from home, or anyone who is away for some reason, such as on holiday or sick.

b Starting with second group of three, taking last value of that group with first value of next group of three.

4 **a** These people have already made the decision to buy and so are unlikely to be representative of people in general.

b This restricts the age range from which the sample can be selected. People outside the small age range of 25 to 29 years do not have a chance of being selected.

c Only men have a chance of being selected.

5 The different totals have different probabilities; not all students have the same chance of being chosen.

6 Number employees from 001 to 712. Generate 3-digit random numbers. Ignore 000 and numbers above 712. Ignore numbers which have come up previously. Repeat the process until 50 distinct employees have been selected.

Exercise 5B

1 $\bar{X} \sim \mathrm{N}\left(6, \frac{8}{80}\right)$; probability = 0.897

2 $\bar{X} \sim \mathrm{N}\left(30, \frac{36}{100}\right)$; probability = 0.0478

3 $\bar{X} \sim \mathrm{N}\left(21, \frac{4.2^2}{50}\right)$; probability = 0.954

4 0.0548

5 0.0019

6 **a** 0.0031 **b** 0.974

7 **a** 0.990 **b** 0.508

8 0.846

End-of-chapter review exercise 5

1 **a** **i** 0.0049 **ii** 0. 932

b Original distribution is normal since n is only 20.

2 **a** $\mathrm{E}(X) = 3.7$

$\mathrm{Var}(X) = 2^2 \times 0.1 + 3^2 \times 0.4 + 4^2 \times 0.2 + 5^2 \times 0.3 - 3.7^2 = 1.01$, as required.

b $S \sim \mathrm{N}(370,\ 101)$

c 0.0207; central limit theorem means S is approximately normal.

3 0.924

4 **a** 0.991 **b** 15.3

5 **a** Not all sums of scores have the same probability.

b Any valid method; random number tables, pieces of paper, Excel function, etc.

Copyright Material - Review Only - Not for Redistribution

6 **i** 59

ii Any value, to three significant figures, from 0.687 to 0.693 inclusive.

iii Repeats may be included, so another random number would have to be chosen.

7 **i** $\bar{N} \sim N(352, 2.9)$ **ii** 0.120

8 **i** Likely to be friends if sitting at the same table; only considers people who use the canteen.

ii 2-digit numbers; ignore > 82; ignore repeats.

9 **i** The sample includes only readers of that issue; it includes only readers who care sufficiently to complete and return the questionnaire.

ii 4975 3952 (0)386 or 4975 5203 6088

10 **i** $N\left(48.8, \frac{15.6^2}{5}\right)$ **ii** 0.568

6 Estimation

Prerequisite knowledge

1 1.5, 0.1, 0.316

2 54.3, 386, 19.6

3 20, 12.9, 3.59

4 13, 23.1, 4.81

5 $H_0: \mu = 86$, $H_1: \mu \neq 86$, $\Phi\left(\frac{84-86}{4}\right) < 5\%$

6 $H_0: \mu = 54$, $H_1: \mu < 54$, $\Phi\left(\frac{50-54}{3}\right) < 5\%$

7 $H_0: \mu = 18$, $H_1: \mu > 18$, $\Phi\left(\frac{20-18}{\sqrt{3}}\right) > 99\%$

8 $X \sim N(16.8, 10.08)$

9 $X \sim N(55, 24.8)$

Exercise 6A

1 $\bar{x} = 2.2$; $s^2 = 0.667$

2 $\bar{t} = 21$; $s^2 = 18.6$

3 $\bar{x} = 4.2$; $s^2 = 0.143$

4 $\bar{x} = 752$; $s^2 = 23.1$

5 $\bar{x} = 8.03$; $s^2 = 7.15$

6 $\bar{x} = 1.1$; $s^2 = 1.55$

7 $\bar{x} = 2.41$; $s^2 = 1.27$

Exercise 6B

1 $H_0: \mu = 18.4$; $H_1: \mu > 18.4$

$$\Phi\left(\frac{19.7-18.4}{\frac{3.6}{\sqrt{40}}}\right)$$

$\Phi(2.284) = 0.9888$

$1.12\% > 1\%$

Accept H_0. There is no evidence to suggest the tablets do not work that quickly.

2 $H_0: \mu = 100$; $H_1: \mu > 100$

a $\bar{x} = 102.4$

$s^2 = 219.4$

$$\Phi\left(\frac{102.4-100}{\sqrt{\frac{219.4}{180}}}\right)$$

$\Phi(2.174) = 0.9852$

$1.48\% < 2\%$

Reject H_0 and accept H_1. There is some evidence to suggest the mean IQ is higher than 100.

b $\bar{x} = 104$

$$\Phi\left(\frac{104-100}{\frac{15}{\sqrt{6}}}\right)$$

$\Phi(0.653) = 0.7432$

$25.68\% > 2\%$

Accept H_0. There is insufficient evidence to suggest the mean IQ is higher than 100.

c Results obtained from a small sample may not be as reliable as those obtained from large samples.

3 $H_0: \mu = 380$; $H_1: \mu < 380$

$$\Phi\left(\frac{378.7-380}{\sqrt{\frac{6.4}{10}}}\right)$$

$\Phi(-1.625) = 0.0521$

$5.21\% > 5\%$

Accept H_0.

Assume variance unchanged. There is insufficient evidence to suggest the mean mass of pesto has reduced

4 $H_0: \mu = 68$; $H_1: \mu < 68$

$\bar{x} = 66.625$

$$\Phi\left(\frac{66.625-68}{\frac{1.7}{\sqrt{8}}}\right)$$

$\Phi(-2.288) = 0.011$

$1.1\% > 1\%$

Accept H_0.

Assume egg masses are normally distributed and variance unchanged. There is no evidence to support the claim that the eggs are underweight.

Copyright Material - Review Only - Not for Redistribution

5 $H_0: \mu = 80$; $H_1: \mu < 80$

$\bar{x} = 77.5$

$$\Phi\left(\frac{77.5 - 80}{\sqrt{\frac{9}{6}}}\right)$$

$\Phi(-2.041) = 0.0207$

$2.07\% < 5\%$

Reject H_0.

Assume variance unchanged. There is some evidence to suggest the amount of ice cream is too low.

6 $H_0: \mu = 2$; $H_1: \mu < 2$

$\bar{x} = 1.97675$

$s^2 = 0.01887$

$$\Phi\left(\frac{1.97675 - 2}{\sqrt{\frac{0.01887}{80}}}\right)$$

$\Phi(-1.514) = 0.065$

$6.5\% > 5\%$

Accept H_0. There is no evidence to suggest the bags of potatoes are underweight.

7 $H_0: \mu = 2000$; $H_1: \mu \neq 2000$

$\bar{x} = 1997.5$

$s^2 = 234.66$

$$\Phi\left(\frac{1997.5 - 2000}{\sqrt{\frac{234.66}{42}}}\right)$$

$\Phi(-1.058) = 0.145$

$14.5\% > 5\%$

Accept H_0.

Assume the lifetime of bulbs follows a normal distribution. There is no evidence to suggest the average lifetime of bulbs is different.

Exercise 6C

1 **a** (1.86, 2.54) **b** (1.96, 2.44)

2 **a** (19.5, 22.5) **b** (19.2, 22.8)

3 **a** (3.88, 4.52, 18.2, 20.1) **b** (18.4, 19.9)

4 **a** (1.94, 2.00) **b** (1.93, 2.02)

5 **a** (7.55, 9.41) **b** 92.3%

6 **a** $\bar{x} = 1.95$; $s^2 = 0.0775$ **b** (1.87, 2.03)

c 90%

7 **a** $2 \times 2.236 = 4.472$ **b** 49

Exercise 6D

1 **a** (0.140, 0.260) **b** (0.115, 0.285)

2 (0.308, 0.428)

3 (0.105, 0.345)

4 **a** (0.507, 0.581) **b** 9

5 45

6 **a** (0.127, 0.273)

b (0.263, 0.437). Since the confidence intervals intersect, from 0.263 to 0.273, at the 99% confidence level it is possible that more than 35% of bees in the colony are infected, and so it might collapse.

End-of-chapter review exercise 6

1 **a** (0.0584, 0.142)

b **i** $0.113 + \frac{1}{2}(0.207 - 0.113)$ or

$0.207 - \frac{1}{2}(0.207 - 0.113)$

ii 80%

2 **a** Average needs to be higher than contents as the sold product cannot be less than that stated.

b This is a small sample so we have to assume that the mass is normally distributed. $H_0: \mu = 276$; $H_1: \mu \neq 276$

$$\Phi\left(\frac{277.7 - 276}{\frac{1.8}{\sqrt{8}}}\right)$$

$\Phi(2.671) = 0.9962$

$99.62\% > 97.5\%$

Reject H_0.

Assume variance unchanged.

3 **a** (19 600, 22 400)

b **i** $1 - 2 \times 0.001 = 99.8\%$ **ii** \$23 194

4 **a** $\bar{x} = 288$; $s^2 = 650$ **b** (283, 293)

5 **a** $\bar{x} = 81.5$

$H_0: \mu = 80$; $H_1: \mu > 80$

$$\Phi\left(\frac{81.5 - 80}{\frac{2.6}{\sqrt{8}}}\right)$$

$\Phi(1.632) = 0.9486$

$94.86\% < 95\%$

Accept H_0. There is insufficient evidence to reject the claim.

Assume sd unchanged. Battery charging times are normally distributed.

Copyright Material - Review Only - Not for Redistribution

b $H_0: \mu = 24$; $H_1: \mu < 24$

$$\Phi\left(\frac{23.2-24}{\frac{1}{\sqrt{5}}}\right)$$

$\Phi(-1.789) = -0.9633$

$3.67\% < 5\%$

Reject H_0.

Assume variance unchanged.

c 26.0 h

6 11, 12, 13, 14 or 15

7 **i** $H_0: \mu = 2.60$

$H_1: \mu > 2.60$

$$\Phi\left(\frac{2.64-2.6}{\frac{0.2}{\sqrt{75}}}\right) = \Phi(1.732)$$

Reject H_0. There is sufficient evidence to suggest the mean has increased.

ii 0.0345

8 **i** One-tailed as the test refers to a possible increase.

ii 1.481 to 1.503 < 1.645

iii Type II possible as H_0 not rejected.

9 **i** $\frac{\bar{x}-22}{\frac{3.5}{\sqrt{12}}} > 1.645$

ii 0.0172

10 **i** (7.54, 9.26)

ii No, since the distribution of the population is normal.

iii 8 g lies within the confidence interval, so the claim is justified.

11 **i** (29.3, 29.9)

ii 30 g lies outside the confidence interval, so the claim is not supported.

iii The confidence interval is a variable.

12 **i** 0.145; $n = 600$ **ii** 97.4

13 **i** There is insufficient evidence to conclude that the mean is less than 750 g.

ii $n = 18$

Cross-topic review exercise 2

1 $\int_1^t \frac{x}{8}\,dx = \int_1^t \frac{9}{4x^3}\,dx$

$\left[\frac{x^2}{16}\right]_0^t = \left[\frac{-9}{8x^2}\right]_1^t$

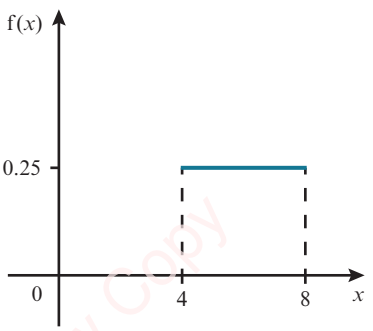
$\frac{t^2}{16} = \frac{-9}{8t^2} + \frac{9}{8}$

$t^4 - 18t^2 + 18 = 0$

$t^2 = 9 \pm 3\sqrt{7}$

$t = 1.03$ is the only possible solution since other solutions outside range for t.

2 **a**



b Mean = 6; variance = $\frac{4}{3}$

c $\bar{X} \sim N\left(6, \frac{1}{30}\right)$

d $a = 6.23$

3 **a** Mean = 14.167; variance = 7.639

b $\bar{T} \sim N(14.167, 0.382)$

c $14.167 - 2 \times \sqrt{0.382} = 12.93 > 12.4$

Alt.

$$\Phi\left(\frac{12.4-14.167}{\sqrt{0.382}}\right) = \Phi(-2.86) = 0.0021$$

4 **i** Testing destroys items; it takes too long or is too expensive to try to test the population.

ii **a** (64.9, 66.5)

b 64.7 lies outside the confidence interval; therefore, the mean height of bounce probably is affected.

5 **i** 2.09; 0.000131 **ii** 0.0832

6 **i** Mean = 30.2; variance = 129

ii Mean = 55.2; variance = 291

7 **i** (0.133, 0.307)

ii Mean = 127; variance = 11 001 or 11 000

iii Generate 4-digit numbers; ignore repeats; ignore values > 9526.

8 **i** 0.146

ii **a** Looking specifically for a decrease.

b There is sufficient evidence to reject H_0; the mean has decreased.

c No, because the population is normally distributed.

9 **i** Mean = 2866 or 2870; variance = 4130

Copyright Material - Review Only - Not for Redistribution

ii There is no evidence to suggest that the mean distance has changed.

10 i (56.6, 67.4)

ii 92% of all confidence intervals will contain the mean.

iii Each possible sample is equally likely to be chosen.

11 i Assume that the standard deviation is unchanged. There is sufficient evidence to suggest that the mean weekly profit has increased.

ii Yes, because the distribution of X is unknown.

iii 0.05

iv 0.0091

12 i Sampling from a telephone directory excludes: people without a phone; those who are ex-directory; children; partners of named person.

ii (0.119, 0.261)

iii $x = 92.8\%$ (2 significant figures)

13 i Mean = 331(.125); variance = 4.125 or 4.13

ii (330, 332)

iii No, because 333 does not lie within the confidence interval.

14 i 0.941

ii $m < 2.53$ and $m > 2.71$

15 i Multiplying by 2 means odd numbers are not included.

ii (1.54, 1.70)

16 i 0.324

ii 0.994

17 H_0: $\mu = 42$; H_1: $\mu \neq 42$. The teacher's estimate can be accepted.

18 i $\mu = 4.27$; $\sigma^2 = 0.00793$

ii (4.25, 4.29)

iii 9

19 i $\int_1^4 kt^{\frac{1}{2}}\,\mathrm{d}t = 1$

$$\left[\frac{2kt^{\frac{3}{2}}}{3}\right]_1^4 = 1$$

$$k = \frac{3}{14}$$

ii 2.66 h iii 2.73 h iv 0.0243

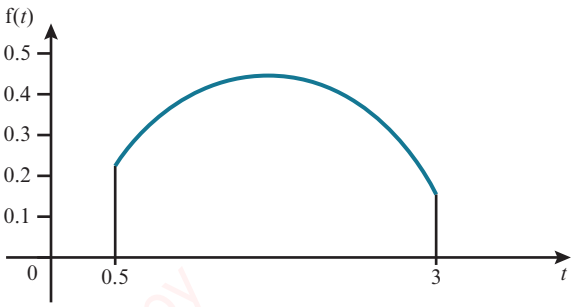
20 i Only train travellers are asked the question, so very many adults are excluded.

ii People travelling to work on this train.

iii Mean = 6.17; variance = 0.657

Practice exam-style paper

1 a



$$k\int_{0.5}^{3} 10t - 3t^2\,\mathrm{d}t = 1$$

$$k\left[5t^2 - t^3\right]_{0.5}^{3} = 1$$

$$k = \frac{8}{135}$$

b i $\frac{8}{135}\left[5t^2 - t^3\right]_2^3 = \frac{48}{135}$

ii $\mathrm{B}\left(10, \frac{48}{135}\right)$

$${}^{10}\mathrm{C}_0\left(\frac{48}{135}\right)^0\left(1-\frac{48}{135}\right)^{10} + {}^{10}\mathrm{C}_1\left(\frac{48}{135}\right)^1\left(1-\frac{48}{135}\right)^9$$

$$= 0.0805$$

c $\frac{8}{135}\int_{0.5}^{3} 10t^2 - 3t^3\,\mathrm{d}t = \frac{8}{135}\left[\frac{10t^3}{3} - \frac{3t^4}{4}\right]_{0.5}^{3} = 1.711$

d Not realistic, as it is unlikely that no calls will be connected within $\frac{1}{2}$ a minute.

2 a i $\mathrm{e}^{-0.9}\frac{0.9^3}{3!} = 0.0494$

ii $0.0494^3 = 0.000121$

iii $\mathrm{e}^{-0.9}\left(1 + 0.9 + \frac{0.9^2}{2}\right) = 0.937$

b $\mathrm{e}^{-(4\times 0.9)} = 0.0273$

c H_0: $\lambda = 3.6$; H_1: $\lambda < 3.6$;

$\mathrm{e}^{-3.6}(1 + 3.6) = 0.126$

As this is greater than 10%, accept H_0. There is insufficient evidence to say that the number of breakdowns has decreased.

3 $\mathrm{B}\left(12, \frac{1}{2}\right)$

$$\binom{12}{0}\frac{1}{2}^{12} + \binom{12}{1}\frac{1}{2}^{12} + \binom{12}{2}\frac{1}{2}^{12} = 0.0193$$

Minimum level of significance = 2%

Copyright Material - Review Only - Not for Redistribution

4 $8 \times 17 + 8 \times 4.5 + 50 = 222$

$8 \times 2^2 + 8 \times 0.3^2 = 32.72$

$\Phi\left(\frac{230 - 222}{\sqrt{32.72}}\right)$

$= \Phi(1.399) = 0.919$

5 **a** $32.2 \pm z \times 3.9$

$z = 1.96$

Required interval $= (24.6, 39.8)$

b $\frac{301.6}{10} = 30.16$

$30.16 \pm z \times \frac{3.9}{\sqrt{10}}$

$z = 1.645$

Required interval $= (28.1, 32.2)$

6 Assume that the lifetime of the batteries is normally distributed.

$H_0: \mu = 76$; $H_1: \mu \neq 76$

$\Phi\left(\frac{74.6 - 76}{\frac{4}{\sqrt{40}}}\right)$

$\Phi(-2.214) = 0.0135$

$1.35\% > 1\%$

Accept H_0; there is insufficient evidence to reject the manufacturer's claim.

Copyright Material - Review Only - Not for Redistribution